

STIC-Biotech/ChemLib

From: Chan, Christina
Sent: Monday, April 12, 2004 10:11 AM
To: Duffy, Patricia; STIC-Biotech/ChemLib
Subject: RE: Please rush amendment due this biweek

Importance: High

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Duffy, Patricia
Sent: Monday, April 12, 2004 8:38 AM
To: Chan, Christina
Subject: Please rush amendment due this biweek
Importance: High

Please approve rush protein sequence below.

In Re: 09/868,195

Please search SEQ ID NO:12. -475 AA

Please include an interference and commercial database search.

Please print out top 50 hits in each category.

Thank you,

Patricia A. Duffy, Ph.D.
Art Unit 1645, Remsen 3B05
571-272-0855

Searcher: D. Schuchter
Phone: 272-2726
Location: Rem 3E89
Date Picked Up: 4/14
Date Completed: 4/14
Searcher Prep/Review: 11
Clerical: 5
Online time: 5

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: CompuLink
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:09:58 ; Search time 17 Seconds
(without alignments)
1454.901 Million cell updates/sec

Title: US-09-868-195-12
Perfect score: 2393
Sequence: 1 LTKYQNYVNGWKSSVQI.....GIRYSIEATNVKSVLDLWK 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1700	71.0	475	1 GAPN_STRMU	Q5931 streptococ
2	1164.5	48.7	498	1 GAPN_MAIZE	Q43272 zea may5 (m
3	1161.5	48.5	496	1 GAPN_PEA	P81406 pisum sativ
4	1155.5	48.3	496	1 GAPN_NICPL	P93338 n nado-depe
5	763	31.9	488	1 DHAI_BACSU	F42236 bacillus su
6	759.5	31.7	463	1 YE11_METJA	Q58806 methanococ
7	726	30.3	482	1 GABD_ECOLI	P25526 escherichia
8	721.5	30.2	488	1 DHAI_BACST	P42329 bacillus st
9	666	27.8	509	1 DHAI_CHICK	P27463 gallus gall
10	660	27.6	483	1 NAHF_PSESP	Q52460 pseudomonas
11	660	27.6	517	1 DHAS_HUMAN	P30837 homo sapien
12	658.5	27.5	489	1 DHAS_ECOLI	P17445 escherichia
13	658	27.5	500	1 DHAI_HUMAN	P00352 homo sapien
14	650	27.2	512	1 DHAI_HUMAN	P47895 homo sapien
15	647	27.0	495	1 DHAI_ECOLI	P23883 escherichia
16	640	26.7	499	1 DHAS_CHICK	Q33344 gallus gall
17	639	26.7	499	1 DHAS_HUMAN	Q94788 homo sapien
18	639	26.7	500	1 DHAI_SHEEP	P51977 ovis aries
19	636	26.6	499	1 DHAI_RAT	Q63639 rattus norv
20	635	26.5	499	1 DHAI_MOUSE	Q62148 mus musculu
21	633	26.5	500	1 DHAI_BOVIN	P48644 bos taurus
22	630.5	26.3	497	1 DHAI_EWENT	P08157 emericeletu
23	630	26.3	500	1 DHAM_MESAU	P81178 mesocricetu
24	625	26.1	519	1 DHAM_RAT	P11884 rattus norv
25	624	26.1	500	1 DHAM_HORSE	P12762 equus cabal
26	623	26.0	500	1 DHAI_MOUSE	P51647 rattus norv
27	622	26.0	500	1 DHAI_MOUSE	P24549 mus musculu
28	622	26.0	519	1 DHAM_MOUSE	P47738 mus musculu
29	621	26.0	517	1 DHAM_HUMAN	P50301 homo sapien
30	620	25.9	501	1 DHAE_MACPR	Q29490 macrosceid
31	618.5	25.8	512	1 Y4UC_RHISN	Q53197 rhizobium s
32	618	25.8	501	1 DHAE_FLEED	Q28399 elephantiulu
33	617.5	25.8	497	1 DHAB_SFIOI	P17202 spinacia ol

ALIGNMENTS

RESULT 1
GAPN_STRMU STANDARD; PRT; 475 AA.

AC Q5931;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.9)
DE (Non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase)
DE (Glyceraldehyde-3-phosphate dehydrogenase [NADP+]) (Triosephosphate dehydrogenase)
GN GAPN OR SMU 676.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG5 / Serotype C;
RX MEDLINE=95270576; PubMed=7751269;
RA Boyd D.A., Cvitkovitch D.G., Hamilton I.R.;
RT "Sequence, expression, and function of the gene for the nonphosphorylating, NADP-dependent glyceraldehyde-3-phosphate dehydrogenase of Streptococcus mutans.";
RT J. Bacteriol. 177:2622-2627(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Important as a means of generating NADPH for biosynthetic reactions.
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + NADP(+) + H(2)O = 3-phospho-D-glycerate + NADPH.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L38521; AAA31091.1; --
CC EMBL; AE014911; AAN58410.1; --
DR

P41751 aspergillus
P13601 rattus norv
Q27640 enchytraeus
P15437 equus cabal
O04895 amaranthus
P42757 atriplex ho
E28237 beta vulgar
Q8CND4 staphylococ
P43503 pseudomonas
P20000 bos taurus
Q25417 leishmania
Q8ROY6 mus musculu
P49189 homo sapien
P46367 saccharomyc
P28037 rattus norv
Q9S795 arabidopsis
O74187 agaricus bi

34 617.5 25.8 497 1 DHAL_ASPNG
35 617 25.8 500 1 DHAC_RAT
36 614 25.7 497 1 DHAL_ENCBU
37 613 25.6 500 1 DHAI_HORSE
38 613 25.6 501 1 DHAB_AMAHP
39 609 25.4 502 1 DHAB_ATRHO
40 608 25.4 500 1 DHAB_BETVU
41 607.5 25.4 514 1 ROCA_STAEP
42 607 25.4 487 1 XYLC_PSEPU
43 607 25.4 520 1 DHAM_BOVIN
44 604.5 25.3 498 1 DHAM_LEITA
45 603.5 25.2 902 1 FIDH_MOUSE
46 599 25.0 493 1 DHAG_HUMAN
47 596.5 24.9 519 1 DHAI_YEAST
48 595.5 24.9 902 1 FTDH_RAT
49 594 24.8 501 1 DHAB_ARATH
50 589.5 24.6 500 1 DHAL_AGABI

```

DR PIR; A57151; A57151.
DR PDB; 1EUH; 22-JUL-99.
DR PDB; 1Q1I; 10-JAN-01.
DR PDB; 1Q16; 10-JAN-01.
DR PDB; 2E0H; 22-JUL-99.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldehyd; 1.
DR PROSITE; PS00070; ALDHDEHYD DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDHDEHYD DEHYDR_GLU; FALSE NEG.
KW Oxidoreductase; NADP; Complete proteome; 3D-structure.
FT NP_BIND 230 235 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 250 250 BY SIMILARITY.
FT ACT_SITE 284 284 BY SIMILARITY.
FT CONFLICT 58 58 S -> A (IN REF. 1).
FT CONFLICT 85 85 V -> I (IN REF. 1).
FT CONFLICT 85 85 V -> I (IN REF. 1).
FT CONFLICT 347 347 A -> T (IN REF. 1).
SQ SEQUENCE 475 AA; 51194 MW; FOA2770AB99552DC CRC64;

Query Match 71.0%; Score 1700; DB 1; Length 475;
Best Local Similarity 66.9%; Pred. No. 3.8e-96;
Matches 318; Conservative 75; Mismatches 82; Indels 0; Gaps 0;

QY 1 LTKYQNYVNGEKSVNQIILSPIDDSLGFVPMTRREVDHMKAGREALPAWALT 60
DB 1 MTKQYNYVNGEKSVNQIILSPIDDSLGFVPMTRREVDHMKAGREALPAWALT 60

QY 61 VVERAQLVKAADIIRDEKETATWLAKESKAYNASVTEVRTADLIRYAAEGIRLST 120
DB 61 YIERAAVLHKVADILNRDKEKIGAVLSKEVAKYSVSEVVRTAEIINYRAEEGLRMEG 120

QY 121 SADGGKMDASTGHKLAVIRQPVGIVLAIPYNYVNLSSKIAPALIGNVVWF 180
DB 121 EVLEGSGFEAAKKKIAVVRREPVGVLVAISFPYVNLSSKIAPALIGNVVWF 180

QY 181 QSSVSGVLAKAFAPAGVNTITGRSGEIGDYIVHEEVEVNFNFTGSPVQRI 240
DB 181 QSSVSGVLAKAFAPAGVNTITGRSGEIGDYIVHEEVEVNFNFTGSPVQRI 240

QY 241 LAGMRPIMLGLGKDAIVLADLNAKQIVAGADYSGQRCTAKRVLVVEVDEL 300
DB 241 MAGMRPIMLGLGKDAIVLADLNAKQIVAGADYSGQRCTAKRVLVVEVDEL 300

QY 301 AKESVNAKLVGDPDFNATVPVDDNSADPISLVVDARQKAGKELNEFKRGRLLT 360
DB 301 VEKIREKVALTIGNPEDDADITPLIDTSADYVEGLINDANDKGAALTEIKREGNLC 360

QY 361 PGLFDHVTLDMLKAWEEPGLPLIRVKDAEVAIAANKSDFGLOSSVFTDRFOKAFDI 420
DB 361 PILFDKVTIMRLKAWEEPGLPLIRVTSVEEIEISNKSSEYGLQASIFTDFFPAPGI 420

QY 421 ANKLEVGTVHNNKTGRGPNFPFLGLKSGAGVQGIYSIEAMTNVKSIVLDMK 475
DB 421 AQLEVGTVHNNKTGRGPNFPFLGLKSGAGVQGIYSIEAMTNVKSIVLDMK 475

RESULT 2
GAPN MAIZE STANDARD; PRT; 498 AA.
AC Q43272;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.9)
DE (Non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase)
DE (Glyceraldehyde-3-phosphate dehydrogenase [NADP+]) (Triosephosphate
DE dehydrogenase).
GN GPNI.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]

```

```

SEQUENCE FROM N.A.
STRAIN=CV_KW5330; TISSUE=Shoot;
RC MEDLINE=94180387; PubMed=7545914;
RA Habenicht A., Hellman U., Cerff R.;
RT "Non-phosphorylating GAPDH of higher plants is a member of the
RT aldehyde dehydrogenase superfamily with no sequence homology to
RT phosphorylating GAPDH."
RL J. Mol. Biol. 237:165-171(1994).
CC !- FUNCTION: Important as a means of generating NADPH for
CC biosynthetic reactions.
CC !- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + NADP(+) + H(2)O
CC = 3-phospho-D-glycerate + NADPH.
CC !- SUBCELLULAR LOCATION: Cytoplasmic.
CC !- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X75326; CAAS3075.1; -.
CC DR PIR; S43833; S43833.
CC DR HSSP; P51977; 1EXS.
CC DR MaizeDB; 78926; -.
CC DR InterPro; IPR002086; Aldehyde_dehydr.
CC DR Pfam; PF00171; aldehyd; 1.
CC DR PROSITE; PS00070; ALDHDEHYD DEHYDR_CYS; 1.
CC DR PROSITE; PS00687; ALDHDEHYD DEHYDR_GLU; 1.
CC KW Oxidoreductase; NADP.
CC FT NP_BIND 247 251 NAD (ADP PART) (BY SIMILARITY).
CC FT ACT_SITE 266 266 BY SIMILARITY.
CC FT ACT_SITE 300 300 BY SIMILARITY.
CC SQ SEQUENCE 498 AA; 53146 MW; 7AF1C0DACA4EE39 CRC64;

Query Match 49.7%; Score 1164.5; DB 1; Length 498;
Best Local Similarity 48.0%; Pred. No. 1.3e-63;
Matches 234; Conservative 88; Mismatches 143; Indels 13; Gaps 5;

QY 5 QYNYVNGEKSVN-QIILSPIDDSLGFVPMTRREVDHMKAGREALPAWALT 62
DB 18 YRYADGGWRTSAGSKVAIVNPTTRKTQYRVOACTQEEVNVKAMDAAKVAQKAWARTPW 77

QY 63 ERAQVILKAAADIERDEKETATWLAKESKAYNASVTEVRTADLIRYAAEGIRLST 122
DB 78 KRADVILKAAALKEHKAPIAECLVKEIAPKADAVSEVRSGLDLSVTAEGRVLGE- 136

QY 123 DEGGKM---DASTG---HKLAVIRQPVGIVLAIPYNYVNLSSKIAPALIGNVVWF 176
DB 137 ---GKLWSDSFPNGERNKYCLSKPLGVVLAIPFPNYPANLAGSKIGPALLIAGNALVL 193

QY 177 KPPTQSGVSLWLAKAFAPAGVNTITGRSGEIGDYIVHEEVEVNFNFTGSPVQ 236
DB 194 KPPTQAVAAALHMHVCHFLAGFPKGLISCVTSGSIGDFLTMHFGVNCISFTGG-DTGI 252

QY 237 RIGKLAMRPIMLGLGKDAIVLADLNAKQIVAGADYSGQRCTAKRVLVVEV 296
DB 253 AISKAGVVPQLMELGGKACIVLEDADLIVSANIVKGGFSYSGQRCTAKRVLVIMESI 312

QY 297 ADLAEKISENVAKLVGDPDFNATVPVDDNSADPISLVVDARQKAGKELNEFKR 356
DB 313 ADAAVQVNAKLAKLVGDPDFDSDITPVVTESANFIEGLVMDAKSKGATFCQYRREG 372

QY 357 RLATGLDHTLDMLKAWEEPGLPLIRVKDAEVAIAANKSDFGLOSSVFTDRFOK 416
DB 373 NLTPVLLDHTVRPDMRLAWEEPGLPLIRVINSVEEIHHCNANFGLGCIPTDINK 432

QY 417 AFDIAKLEVGTVHNNKTGRGPNFPFLGLKSGAGVQGIYSIEAMTNVKSIVLDM 474
DB 433 ALLISDMETGTVQINSAPARGDFHFSFQGLKDSGIGSGQITNSINMTVKVKSTVINL 490

```

RESULT 3
GAPN_PEA
ID GAPN_PEA STANDARD; PRT; 496 AA.
AC P81406;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.9)
DE (Non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase)
DE (Glyceraldehyde-3-phosphate dehydrogenase [NADP+]) (Triosephosphate
DE dehydrogenase).
GN GAPN.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=cv. Rosakrone; TISSUE=Shoot;
RA MEDIUM=94180387; PubMed=7545914;
RX Habenicht A., Hellman U., Cerff R.;
RT "Non-phosphorylating GAPDH of higher plants is a member of the
RT aldehyde dehydrogenase superfamily with no sequence homology to
RT phosphorylating GAPDH.";
RL J. Mol. Biol. 237:165-171(1994).
CC -!- FUNCTION: Important as a means of generating NADPH for
CC biosynthetic reactions.
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + NADP(+) + H(2)O
CC = 3-phospho-D-glycerate + NADPH.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
DR EMBL; X75327; CAA53076.1; --
DR HSRP; P51977; 1XS
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldesh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NADP.
FT NP BIND 245 249 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 264 264 BY SIMILARITY.
FT ACT_SITE 298 298 BY SIMILARITY.
SQ SEQUENCE 496 AA; 53004 MW; 52FADC2F77C613FA CRC64;
Query Match 48.5%; Score 1161.5; DB 1; Length 496;
Best Local Similarity 49.5%; Pred. No. 1.9e-63;
Matches 236; Conservative 85; Mismatches 145; Indels 11; Gaps 5;
QY 5 YQNYYNGEWNKSSVN--QIEILLSPIDDSSGLGFVPAMTREEDVHAMKGREALPAWAALTIVY 62
DB :
Db 16 KYKADHGKKSTSGKSVALINPTRPKPVKVQAQCSQEENVKMDSAKSKSWAKPLW 75
QY 63 ERAQLHKAADIETTERKEIATVLAKESIKAYNASVTVEVRTADLTIRYAEEGRILSISA 122
: :
Db 76 KRAELHKAALIKEHAALAECLVDIEAPKAKAVTEVRSGDLSVCAEBGVRI---L 132
QY 123 DEGGKM--DASTGH---KLAVTRQPVGIVLAIPNPYPNVISGSKIAPALIGNNVMFK 177
: :
Db 133 GEKGFVSDSPFGNETRKYCTSLPIGLVILAIPFPYPNVAISKVAPALIAGSIVLK 192
QY 178 PPTQGSVGLVIKAFAEAGLPAGVFNTITRGSGSIGDYTVHEEVNFINTGSTPEVQR 237
Db 193 PPTQGAVALHWCHFHLAGPPKGLISCVTGKSGSIGDFLTHHPGVNCISFTGG-DTGIA 251

QY	5	YQNYVNGEWSVNV--QIEILSPDDSSGLGVPAMTREEDVHAKGAREALPANAALTVY	62
Db	16	FKYISEGEWKSASGKSVAILNPTTRTKYQVKQACTQEVNVKNVMEVAKTAQKSWAKTFLW	75
QY	63	ERAQYLHKAADITIERKEBIATVLAKEISKAYNASVTEVETADLTRYAEEGRILSTSA	122
Db	76	KGAELHKAALILKEHAKPIAECHLVKSIAPKADPAVTVRSNGDLVSYTAEEGVRI---	132
QY	123	DEGGKM--DASTGH--KLAVIRROPVGVILATPNYVPNLSGSIAPALIGNVVWFK	177
Db	133	GEKFIIVSDSPGNETKYCLTSKPILOVILAPFNIPVNLVASKIAPALLAGNSLVLK	192
QY	178	PPTQGSVGLVLAFAFAEAGLPAGVFNITITGRGSEIGDYIVIEHEEVNFINTFGSTPVGQR	237
Db	193	PPTQGAVALHWHCHPLAGFEKGLISCVTKGSEIGDFLTMHPGVHCISFTCG-DTGVA	251
QY	238	IGKLAGMPIMLELGGKDGAGIVLADALDNNAAQIVAGAYDISQRCITAKVLVWEVA	297
Db	252	ISKAGAMPIQWELGGKDACIVLEDADLDLAAGSIVKGGFSYSGQRCITAKVVLVWESVA	311
QY	298	DELAEEKISENVAKLSVCGDPDPAVTPVVIDNSADFIESLVVDARQKAGKELNEFKDGR	357
Db	312	DALVEKVNAAKAVLUTGPPEDDCTIPVWESSANFIEGLVNDAKQXNATFCQYKREGN	371
QY	358	LLTPGLFDHVLDMKLAEEPPGPIPLTIIRVKDAEAAVAIAANKSDFGLOSSYFTDFQKA	417
Db	372	LIWPLLDNVPRMIAEPPGPIPVIRINSVESGIHHCNAGSNFGLQGCVFTKDINKA	431
QY	418	FDTANKLEVGTVHNNKTRGDDNPFFLUGLKGSGAGVQGIIRYSIEATNVKSVILDM	474
Db	432	ILISDAMETGTVQINSAPARGDPHPFGQIGKDSGISQGSITNSINMMTKYKTVINL	488

RESULT 5

ID	DHAI_BACSU	STANDARD;	PRT;	488	AA.
AC	P42236;				
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Probable aldehyde dehydrogenase ycbd	(EC 1.2.1.3).			
OS	YCBD OR BSU02470.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]	SEQUENCE FROM N.A.			
RP	STRAIN=168;				
RC	MEDLINE=95219079; PubMed=7704254;				
RA	Qgawa K.-I., Akagawa E., Nakamura K., Yamane K.;				
RT	"Determination of a 21548 bp nucleotide sequence around the 24				
RL	degrees region of the Bacillus subtilis chromosome."				
RL	Microbiology 141:269-275 (1995).				
RN	[2]	SEQUENCE FROM N.A.			
RP	STRAIN=168;				
RC	MEDLINE=98044033; PubMed=9384377;				
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,				
RA	Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,				
RA	Brouillet S., Bruschii C.V., Caldwell B., Capuano V., Carter N.M.,				
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,				
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,				
RA	Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,				
RA	Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,				
RA	Ghim S.Y., Goffeau A., Goffeau E., Goughly E.J., Grandi G.,				
RA	Guileppi G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,				
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,				
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,				
RA	Kurita K., Lapidus A., Lardinios S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,				

RESULT 5

КЕНОТ 3
ДНА1 ВАСИИ

DEAL - BACSU
TD - DIA - P

ID	NAME	AGE	SEX	REL	STATUS	DATE	TIME	LOCATION	REMARKS
1	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
2	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
3	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
4	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
5	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
6	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
7	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
8	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
9	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
10	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
11	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
12	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
13	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
14	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
15	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
16	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
17	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
18	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
19	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
20	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
21	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
22	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
23	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
24	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
25	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
26	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
27	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
28	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
29	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
30	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
31	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
32	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
33	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
34	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
35	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
36	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
37	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
38	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
39	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
40	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
41	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
42	ADAM	15	M	SON	PRESENT	10/10/2023	10:00	HOME	HEALTHY
43	ADAM	15	M	SON	PRESENT	10/10/2023			

AC P42236

DT 01-NOV

DT 01-NOV

DT 10-OC

DE Probab

DE
GN
YCBT
FLOBA

GN
CC
ICB
P

OS Bacilli

OC Bacter

OX NCBI_

RN [1]

RP SEQUEN

RC STRAIN

RC
PY

RA MEDLIT
PA Oaswa

RA
OGAWA
RERT
"Deter

RT degree

RL Microf

RN [2]

RP SEQUENCING

RC
STRAT

RC
PY

RX	MEDLI
P3	Y2-08-06

RA Kunst

RA Azevedo

RA Borri

RA Broui

RA Choi

RA Deniz.

RA
Dell
EntiaRA
Encl
Encl
Encl

RA Fritz

Db 365 KYONGYVQPAIFDNTVMTAQEITFGPVALIKVDSIEBALNIANDVKGLSASIT 424
 QY 412 RFOQAFDIAKLEVTGTHNNKTRGPDNFFPLGLKSGA 452
 Db 425 ENIGMJSFIDEIDAGLRINAESAGVLOAPFGMKOSS 465

RESULT 6

YELL_METJA
 ID YELL_METJA STANDARD; PRT; 463 AA.
 AC Q58806;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical aldehyde-dehydrogenase like protein MJ1411 (EC 1.2.1.-).
 GN MJ1411.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow P.W., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Hannon M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67581; AAB99418.1; -;
 DR PIR; B64476; B64476.
 DR HSSP; P51977; 1BXS.
 DR TIGR; MJ1411; -;
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldehyd; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; FALSE_NEG.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; FALSE_NEG.
 KW Hypothetical protein; oxidoreductase; complete proteome.
 FT ACT SITE 240 274 BY SIMILARITY.
 FT ACT SITE 274 274 BY SIMILARITY.
 SQ SEQUENCE 463 AA; 51135 MW; 467265686595B6A5 CRC64;

Query Match 31.7%; Score 759.5; DB 1; Length 463;
 Best Local Similarity 37.1%; Pred. No. 4.4e-39;
 Matches 175; Conservative 95; Mismatches 183; Indels 19; Gaps 8;

QY 8 YVNGEKSSVNO--TEILSPIDDSIGFVPAMTREVVDHAKAGREALPAWALTIVYRA 65
 Db 2 FIDGKA---INRDMVINYSLEVIKKIPALSRKAEIATAEYKEVWKNLPITKY 58
 QY 66 QYLHKAADIIERKEEATYLAKEISKAYNASVTEVTRADILIRYAEEGIRLSTGADG 125
 Db 59 NILMNTAKQKEKEELAKILAIDAGKPIQARVERSGTGFKLA---FYVKEHRDEV 115
 QY 126 GKMDASTGHLAVRPGVIGVILATAPYNPVNLGSKTAPALIGNVWEKPTGGSV 185
 Db 116 IPSD----DRLIFTRREFVGIVGALTFFNFFNLNSAHKIAPATATGNVIVHPSPKAPLV 171

QY 186 GLVIK---AFABAGLPAGVNTITGRSGEIGDYIVEHEEVNFINFTGSPVGORIGKL 241
 Db 172 CIELAKIENALKKYNVPLGVNLTGAGEVVGDEIVNEKNMISFTGSSKVGELLITKK 231
 QY 242 AGMRPIMLELGKDGAGIVLADADLNAAKQIVAGANDYSGQRCTAKRIVVVEVADELA 301
 Db 232 AGFKIALELGGVNPVILKADLANKAVNALIKGSFTYAGQVCISVGMILVDESIAKFI 291
 QY 302 EKISENVAKLSVGDPFNAT-VTPVIDNSADFTESLVVDARQKGAKEINFEKDGRLIT 360
 Db 292 EMFVNKAKVLNVNPLDEKTDVGLLSVEHAWEKVEKKAIDEGKLLGLGKRAKALFY 351
 QY 361 PGLFDHVTLDMLKAWPEPFGPIPIIRVKDAEAAVAITANKSDFLQSSVFTDRPQAFDI 420
 Db 352 PTLILE-VDRDNILCKTETFAPIPIRTNE-EEMDIANSTEVGLHSAIFTNDINKSLKF 409
 QY 421 ANKLEVTGTHNNKTRGPDNFFPLGLKSGAGVQGRIRYSIEMTNVISVL 472
 Db 410 AENLEFGGVINDSLSLFRQDNMPFGVKKSLGREGVKYKAMEEMSNIKITII 461

RESULT 7

GABD_ECOLI
 ID GABD_ECOLI STANDARD; PRT; 482 AA.
 AC P25526; P78207; P78208; P78209;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16) (SSDH).
 GN GABD OR B2661.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / JM103;
 RX MEDLINE=94127927; PubMed=8297211; K;
 RA Niegemann E., Schulz A., Bartsch K.;
 RT "Molecular organization of the Escherichia coli gab cluster:
 RT nucleotide sequence of the structural genes gabD and gabP and
 RT expression of the GABA permease gene.";
 RL Arch. Microbiol. 160:454-460(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-Kb sequence of the Escherichia coli
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC -!- CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(P)(+) + H(2)O =
 CC succinate + NAD(P)H.
 CC -!- PATHWAY: 4-aminobutyrate (GABA) degradation.
 CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts

```

CC      in positions 51 and 190.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M88334; AAC36831.1; -.
CC      EMBL; AE000351; AAC75708.1; -.
CC      EMBL; D90890; BAA16522.1; ALT FRAME.
CC      EMBL; D90890; BAA16523.1; ALT FRAME.
CC      EMBL; D90890; BAA16524.1; ALT FRAME.
CC      EMBL; F65045; F65045.
CC      PIR; F65045; F65045.
CC      HSP; F51977; JBX5.
CC      EcoGene; EG11329; gabb.
CC      DR InterPro; IPR002086; Aldehyde_dehydr.
CC      Pfam; PF00171; aldedh; 1.
CC      PROSITE; PS00687; ALDEHYDE_DEHYDR_CYS; 1.
CC      PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC      Oxidoreductase; NADP; Complete proteome.
CC      NP BIND 233 238 NAD(P) (ADP PART) (BY SIMILARITY).
CC      FT ACT SITE 255 255 BY SIMILARITY.
CC      FT ACT SITE 289 289 BY SIMILARITY.
CC      SQ SEQUENCE 482 AA; 51720 MW; 091538F8741DB0CF CRC64;

Query Match      30.3%; Score 726; DB 1; Length 482;
Best Local Similarity 34.3%; Pred. No. 5e-37;
Matches 164; Conservative 92; Mismatches 206; Indels 16; Gaps 7;

CC      6 QNYVNGKSSVN--QIEILSPIDDSGLGFPVPMTRFVDHAKGREALPAAALTYVE 63
CC      12 QALLNGELDANGAIDVTPANGDKLSGKPKGADETRAAIDAANRALPAAALTYAKE 71
CC      64 RAQYLHKAADIIRKDEIATVLAKEISKAYNASVTEVVTADLIRYAEIGIRLSTAD 123
CC      72 RATTIRWFNLMWHEQDDRLARLTLEQKPLAEAKGEISYAASFIWFAESGRYIGDTI 131
CC      124 EGGKMDASTCHKLAVIRQPGVILATAPYNPVNLSSGIAPALIGNVVMFKPPTQGS 183
CC      132 FGHQAD-----KRLIVIKOPIGVTAITPNFPAAKITRAGPALAAGCTWVLKPSQTP 186
CC      184 VSLGLVLAFAFAGLPAAGVNTITGRSGEIGDIYVEHEEVNFINFTGSPVQGRIGLAG 243
CC      187 FSALALAEALIRAGVPAGVNVVTSAGAVGNELFNLVKLSFTGSTEIGRLMEQCA 246
CC      244 --MRPIMELGGKDGAGIVLADADLDNAKQIVAGAYDSQGRCTAKRVLVVEVDELA 301
CC      247 KDIIKVSLELGGNAPFIVFDDADLDKAVEGALASKFRNAGQTCVCANRLYVQDGYDRFA 306
CC      302 EKISENVAKLSVGDPPDN-ATVTPVIDDSDAFIESLVVDAROKGAKEL--NEFKRDR 357
CC      307 EKIQAQVSKLHIGDGLDNGVITGLIDERAKAKVEHIAADALEKARVVCCKAHERG 366
CC      358 LITFGLFDHVLDMKLAMEEPPFPIPIIRVKDAEVAIAJANKSDFGLOSFTFRDQKA 417
CC      367 PFQPTILVDPANAKVSEETFGPLAPREFKDEADVIAQANDTEFGLAAFYARDLSRV 426
CC      418 FDIANKLEVTVHINKTG-RGPONFFFLGLKSGAGVQGRYRISIEAMTVKLSIVLDM 474
CC      427 FRVGEALEYIGVGIN--TGITISNEVAPFGIKASGLREGSKYGIEDYLEIKYMCIGL 482

RESULT 8
DHAL_BACST      STANDARD;      PRT; 488 AA.
AC      P42329;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Aldehyde dehydrogenase, Thermotable (EC 1.2.1.3).

```

```

GN      ALDHT.
OS      Bacillus stearothermophilus.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX      NCBI_TaxID=1422;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-SICI;
RA      Imanaka T., Ohta T., Sakoda H., Widyastuti N., Matsuoka M.;
RT      "Cloning, nucleotide sequence, and efficient expression of the gene
RT      coding for thermostable aldehyde dehydrogenase from Bacillus
RT      stearothermophilus, and characterization of the enzyme.";
RL      J. Ferment. Bioeng. 76:161-167(1993)
CC      -!- FUNCTION: Oxidizes several aliphatic aldehydes, particularly c6-
CC      aliphatic aldehyde and hexanal, but do not oxidize benzaldehyde.
CC      -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC      -!- COFACTOR: Requires either NAD or NADP as a coenzyme.
CC      -!- PATHWAY: Ethanol utilization; second step.
CC      -!- SUBUNIT: Homotetramer.
CC      -!- MISCELLANEOUS: The optimum temperature is around 55 and 60 degrees
CC      Celsius.
CC      -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; D13846; BAA02975.1; -.
CC      PIR; I39769; I39769.
CC      HSP; P51977; LBXS.
CC      InterPro; IPR002086; Aldehyde_dehydr.
CC      Pfam; PF00171; aldedh; 1.
CC      PROSITE; PS00687; ALDEHYDE_DEHYDR_CYS; 1.
CC      PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC      Oxidoreductase; NAD.
CC      NP BIND 233 238 NAD (ADP PART) (BY SIMILARITY).
CC      FT ACT SITE 255 255 BY SIMILARITY.
CC      FT ACT SITE 289 289 BY SIMILARITY.
CC      SQ SEQUENCE 488 AA; 52915 MW; 29E824451985D9ED CRC64;

Query Match      30.2%; Score 721.5; DB 1; Length 488;
Best Local Similarity 34.4%; Pred. No. 9.4e-37;
Matches 166; Conservative 104; Mismatches 192; Indels 21; Gaps 9;

CC      3 KEYQVYNGKSSV-NOIE-ILSPIDDSGL-GLFVPMTRFVDHAKGREALPAAAL 59
CC      8 KYTFNYINGNVSVSVNVSINPANRHDIVGYVQSTLEDVNEAVTAENAQAQISWKR 67
CC      60 TYPERAQYLHKAADIIRKDEIATVLAKEISKAYNASVTEVVTADLIRYAEIGIRLS 119
CC      68 SGVERGEYLYKAAHILEQCLDIAETMTREMGKTLAEAKAETMRGVHILRYAGEGAR-- 125
CC      120 TSADSGGMDASTCHKLAVIRQPGVILATAPYNPVNLSSGIAPALIGNVVMFKXP 179
CC      126 ---KTGDIVPSSSEGLLFTTRKVLGVGVISPNFVVAIPINMAPALVYNTVLPKA 182
CC      180 TQGSVSLGLVLAFAFAGLPAAGVNTITGRSGEIGDIYVEHEEVNFINFTGSPVQGRIG 239
CC      183 SETATVAAKVIECFHEAGFPKGVNVVCGSVVGGIANGHPDIDGVTFTGNTVKGQVG 242
CC      240 KIAGMR--PIMELGGKDGAGIVLADADLDNAKQIVAGAYDSQGRCTAKRVLVVEVA 297
CC      243 RAAFERGAKYQLEWGGKNPVIVAKDADLDLAVEGTSIGSLRSTQCKTATSRVTEREVY 302
CC      298 DELAEKISENVAKLSVGDPPDNAT-VTPVIDDSDAFIESLVVDAROKGAKELNEFKR-- 354
CC      303 EPPKALLERVKQLKNGLDLAEATWVGFCASESQFHTVLVSIEKSGEAKLIYGNRL 362
CC      355 -----DGLSLTPGLFDHVLDMKLAMEEPPFPIPIIRVKDAEVAIAJANKSDFGLOS 409

```

```

Db 363 EGELANGFEVEPTIFEDVLOMTIAREIFGPVLALIQVDSIEBAIKLANDTEYGLSASI 422
QY 410 FTRDQKAFDIANKLEVGTVHNNKTGRGPNDFPGLK--GSGAGVQGIYSIEAMTNV 467
Db 423 YTKNIGNALEFKIDIEAGLIKVNATAGVFOAPFGMKQSSHSREQG--QAAIEFFTSI 481
QY 468 KSI 470
Db 482 KIV 484

RESULT 9
DHAI CHICK STANDARD; PRT; 509 AA.
AC P2743;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1).
GN ALDH1A1
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic retina;
RX MEDLINE=92217647; PubMed=1559558;
RA Godbout R.;
RT "High levels of aldehyde dehydrogenase transcripts in the
RT undifferentiated chick retina.";
RL Exp. Eye Res. 54:297-305(1992).
CC -!- FUNCTION: Binds free retinal and cellular retinol-binding protein-
CC bound retinal. Can convert/oxidize retinaldehyde to retinoic acid
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SUBUNIT: Homotrimer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X58869; CA941679.1; --
DR F1R; S14629; S14629.
DR HSP; P51977; IBSX.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 254 259 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 277 277 BY SIMILARITY.
FT ACT_SITE 311 311 BY SIMILARITY.
SQ SEQUENCE 509 AA; 55809 MW; 7771181FA2F05DA9 CRC64;

Query Match 27.8%; Score 666; DB 1; Length 509;
Best Local Similarity 33.9%; Pred. No. 2.3e-33;
Matches 163; Conservative 86; Mismatches 204; Indels 26; Gaps 9;

QY 8 YVNGKSSV--QIEILSPIDSSLGFPVPAITREEDVDMKAGREAL---PAWAALTVY 62
Db 32 FINNEHDSVSGKEFVFNANEKICEVAGDADIDKAVKAKKAFELGSPNFDAS 91
QY 63 ERAQYLKAAADIIEEDKEEIIATVIA-----KSIKAYNASVTEVTAIDRYAERGIRL 118

```

```

Db ERGRLEKLNKLADI.VERDRULTLATMEADGGKLFSTAYLMDLGACIKT---IRYCAWADKI 148
QY 119 STADGGGMDASTGHKLAVIRROQVIGVLAIPYPPVNLSSKXIAPALIGNVVMFKP 178
Db 149 H-----GRTPMDGNFTFTRHPFVGCGQIIPWNPFWFIWKIAPALCCGNTVVVKP 202
QY 179 PTGSGSVGLVAKAFABAGLPAGVFNITGRSGEIGDIYIVHEHEVNFNFTGSTPPVQRI 238
Db 203 AEQTPLSALYMGSLIKEAGFPVGVNIVPGPTAGAAISHHMDIDKVSFTSGTEVGKLI 262
QY 239 GKLAG---MRPIMLELGGKAGIVLADADLNAAQIVAGDYSGQRTAKIKVLWVEE 295
Db 263 KEAAGKTNLKRVTLELGKSPNIIFADADLDEAAEFAHGLFYHQGCCICAGSRIFVEEP 322
QY 296 VADELAEKISENVAKLSVGPDP--FDNATVTPVIDNSADFTESLWVDAKQAK---ELNE 351
Db 323 IYDEFVRRSIERAKKTYLGDPLLPVQVQGPQIDKQKILDLIESGKKEGAKLECGGP 382
QY 352 FKRDGRLTLPGLFDHVTLDKMLAWEPFGPILPIRVKDAEAVAIAANKSDFGLQSSVFT 411
Db 383 WGNKGFIQPTVFSNVTDMDRIAKEEIFGVPQIQMKFKTIDEVIKRNNTTYGLAAAVFT 442
QY 412 RDKQKAFDIANKLEVGTVHNNKTGRGPNDFPGLKSGAGVQGITRYSIEAMTNKSI 471
Db 443 KDIDKALTFSALQAGTVWVNCYSAFSA--QCPFGGFRMGNGRELGEYGLQEYTEVKITV 501
QY 472 L 472
Db 502 I 502

RESULT 10
NAHF PSESP STANDARD; PRT; 483 AA.
AC Q52460; O34269;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salicylaldehyde dehydrogenase (EC 1.2.1.65).
GN NAHF OR DOXF.
OS Pseudomonas sp. (strain C18), and
OS Pseudomonas putida.
OG Plasmid NP11, and Plasmid NPL-41.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306, 303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C18;
RX MEDLINE=94042852; PubMed=8226631;
RA Denome S.A., Stanley D.C., Olson E.S., Young K.D.;
RT "Metabolism of dibenzothiophene and naphthalene in Pseudomonas
RT strains: complete DNA sequence of an upper naphthalene catabolic
RT pathway.";
RL J. Bacteriol. 175:6890-6901(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=P. putida; STRAIN=BS202; PLASMID=NPL1;
RA Bezborodnikov S.G., Boronin A.M., Tiedje J.M.;
RT "Nucleotide sequences of genes encoding an upper pathway of
RT naphthalene metabolism of NPL1 plasmid from Pseudomonas putida strain
RT BS202.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 221-483 FROM N.A.
RC SPECIES=P. putida; STRAIN=BS202; PLASMID=NPL-41;
RA Piruzian E.S., Serebryskaya T.S., Lenets A.A., Goldenkova I.V.,
RA Kobets N.S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Salicylaldehyde + NAD(+) + H(2)O = salicylate
CC + NADH.
CC -!- PATHWAY: Upper naphthalene catabolic pathway which involves
CC conversion of naphthalene to salicylate, and catabolism of

```

CC dibenzothioephene (DBT) and phenanthrene. Converts phenanthrene to
CC 1-hydroxy-2-naphthoic acid and the metabolism of DBT is limited to
CC oxidation of the aromatic ring.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; M60405; AAA16129.1; -.
DR EMBL; AF010471; AAB62710.1; -.
DR EMBL; Y14173; CAA74578.1; -.
DR HSSP; P51977; 1BX5.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; FALSE NEG.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
FT NP_BIND 228 233 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 250 250 BY SIMILARITY.
FT ACT_SITE 284 284 BY SIMILARITY.
FT CONFLICT 318 318 V -> RL (IN REF. 3).
FT CONFLICT 365 365 I -> M (IN REF. 3).
SQ SEQUENCE 483 AA; 51994 MW; CD606BADD21DFACE CRC64;

Query Match 27.6%; Score 660; DB 1; Length 483;
Best Local Similarity 34.9%; Pred. No. 5e-33;
Matches 166; Conservative 79; Mismatches 216; Indels 14; Gaps 6;

QY 8 YVNGEKKSSVQ--TEILSPIDSSLGFPVPMTRTEVDHAKAGREALPAWALTYYERA 65
DB 6 FINNMTDSSDQQTERTHPVSSDVVTESANATVDAIKAAQABEAFTKWAQPSERR 65

QY 66 QYLHKAADIIRKEETATVLAKEISKAYNASVTEVVRTADLIRVAEEGRILSTADSG 125
DB 66 RLLLVKQVWESKTPKFTVNAVEVGASALWAGENVHASANVPREAS--LATQI-QG 120

QY 126 GKMDASTGHLAVTERQPVGVLTALAPYNPVNLGSKIPALIGNVVMEKPTQGSVS 185
DB 121 ETPTDKAETLSMTURQVGPDLISVFWNGTAVLAARAIYPLVCGNVTFVSGSEFSPAT 180

QY 186 GLVLAKAFAGLIPAGVNTITG---RGSEIGDYIVEHEEVNFINTGSTVPVGRIKLA 242
DB 181 HALITQVQVQVAGLIPAGVLNLSNPPDRSPEIADALISAKEIRINFTGSTRVSGIIAQKA 240

QY 243 G--WRPIMLELGGKAGIVLADADLNAKQIVAGADYSGQCTAIKRVLVVEVADEL 300
DB 241 AQHLKRLLELGGKSPFLVLDADADDAVAAVAFSGELFQGLQICMSTERLIVDEKIADEF 300

QY 301 AEKISENVAKLSDPP--FDNATVPTVIDNSADPESLVVDARQKGAKELEKFRDGL 358
DB 301 VAKVEVTKLSDGDPCTGDCILGPMVSPNNGERINGLFKDIDKAGKVCVGGIAGAL 360

QY 359 LTPGLFDHVTLOKLAWEEPGLPIPIRVKDAEVAIAKSDFGLOSSVFTDFQKAF 418
DB 361 MPATILDHVKSDMRIYDEETGPITVTVIRCKGEAVRIANDSVGLSGSVGFRDINRAL 420

QY 419 DIANKLEVTVHINKTGQDPNFPFLGLKSGAGVQGIYRSTYEMTNVKSIVLD 473
DB 421 RVGMSIEYGVSHINGSTVQNEAQAPYGGTKNTGYGFRDGRAVIDEFTIKMLTIE 475

RESULT 11
DHAS_HUMAN STANDARD; PRT; 517 AA.
AC P30837;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Aldehyde dehydrogenase X, mitochondrial precursor (EC 1.2.1.3) (ALDH
DE Class 2).
GN ALDH1B1 OR ALDH5 OR ALDHX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=91286241; PubMed=2061311;
RA Hsu L.C., Chang W.-C.;
RT "Cloning and characterization of a new functional human aldehyde
RT dehydrogenase gene.";
RL J. Biol. Chem. 266:12257-12265 (1991).
[2]
RN VARIANTS VAL-86 AND ARG-107.
RX MEDLINE=94063859; PubMed=8244338;
RA Sherman D., Dave V., Hsu L.C., Peters T.J., Yoshida A.;
RT "Diverse polymorphism within a short coding region of the human
RT aldehyde dehydrogenase-5 (ALDH5) gene.";
RL Hum. Genet. 92:477-480 (1993).
CC -1- FUNCTION: ALDHs play a major role in the detoxification of
CC alcohol-derived acetaldehyde. They are involved in the metabolism
CC of corticosteroids, biogenic amines, neurotransmitters, and lipid
CC peroxidation.
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -1- SUBUNIT: Ethanol utilization; second step.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- TISSUE SPECIFICITY: Liver, testis and to a lesser extent in brain.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; M63967; AAA96830.1; -.
DR HSSP; P20000; 1AG8.
DR Genew; HGNC:407; ALDH1B1.
DR MIM; 100670; .
DR InterPro; IPR002085; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism.
FT TRANSIT 1 17 MITOCHONDRION (POTENTIAL).
FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE X.
FT NP_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 285 285 POTENTIAL.
FT ACT_SITE 319 319 POTENTIAL.
FT VARIANT 86 86 A -> V (in allele ALDH1B1*2).
FT VARIANT 107 107 L -> R (in allele ALDH1B1*3).
FT VARIANT 107 107 L -> R (in allele ALDH1B1*3).
SQ SEQUENCE 517 AA; 57217 MW; 78AC173AAC687165 CRC64;

Query Match 27.6%; Score 660; DB 1; Length 517;
Best Local Similarity 34.8%; Pred. No. 5.5e-33;
Matches 170; Conservative 80; Mismatches 201; Indels 38; Gaps 11;

QY 6 QNYVNGEKKSSVQ--TEILSPIDSSLGFPVPMTRTEVDHAKAGREAL---PAWALT 60
DB 38 QLFINNEWQDQVSKTFTPTVNTTGEVIGHVAEGDRAVDRAVKAAREAFRLGSPWRMD 97

QY 61 VYERQVILHKAADIIRKEETATV----LAKESKAYNASVTEVVRTADLIRVAEEGI 116
DB 98 ASERGRLLNLADLVERDVRVIALETLDNGKPFQESYALDLDDEVK---VYRYFA--- 150

```

QY 117 RLSTSADE-GKKMDASTGHKLAVIRROPVIGVILAIAPNYPVNLSSGIAPALIGNVVM 175
Db 151 ---GWADKXHGKTIPEHGOHCFTHFHPVGVCGIIPNPNFLVWGWKLAPALATGNTIV 207
QY 176 FKPPQGSVGLVLAKAFARAGLPAGVFNITGRSGEIGDVIVHEEVNFINTGSTPVG 235
Db 208 MKVABQTPLSALVILASLIKAEFGPPGVNITIGYGTAGAAIAQHMDVDKVAFTGSTVG 267
QY 236 QRIKLAG--MRPIMLEGGKAGIVLADADLNNAAQIVAGAYDYSGQRTAKRVLV 292
Db 268 HUIQKAAGDSNJKRVTLVGGKSPSIVLADADMEHAEVQCHALFFNMGGCCAGSRFTV 327
QY 293 VEEVADLAELAKISENVAKLVSQDPFDNATVT-PVIDDINSADFIESLVVDAROKGAKEL-- 349
Db 328 EESIYNELETVKAKQKQKVGNPPELDTQGGQVQDKQEPFVLVGLQKQEGAKLIG 387
QY 350 -NEFKRDGRLLTPGLFDHVTLMKLMWEPFPIPIRIVKDAEAVAIANKSDFGLOSS 408
Db 388 GERFERGFFTKPTVFGVQDDMRIAKEIFGVPQPLFKFKIEVVERANNTRYGLAAA 447
QY 409 VETROFQKAFDIANKLEVGTGHHNKTGRGPDNF-----PFLGLKSGSAGVQGIYSIEA 463
Db 448 VTRDLDKAMVFTQALQAGTVWNTY-----NIVTCHTPGGFKESGNGRELGDGKA 501
QY 464 MTNVKSIVL 472
Db 502 YTEVKTVTI 510

RESULT 12
ID DHAB_ECOLI STANDARD; PRT; 489 AA.
AC P1745;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 15, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Betaine aldehyde dehydrogenase (EC 1.2.1.8) (BADH).
GN BETB OR B0312.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 1-28.
RP STRAIN=K12;
RC MEDLINE=92065800; PubMed=19562895;
RA Lankark T., Kaasen E., Eshoo M.W., Falkenberg P., McDougall J.,
RA Strom A.R.;
RT "DNA sequence and analysis of the bet genes encoding the
RT osmoregulatory choline-glycine betaine pathway of Escherichia coli.";
RL Mol. Microbiol. 5:1049-1064(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91348527; PubMed=1879697;
RA Boyd L.A., Adam L., Pelcher L.E., McHughen A., Hirt J., Selvaraj G.;
RT "Characterization of an Escherichia coli gene encoding betaine
RT aldehyde dehydrogenase (BADH): structural similarity to mammalian
RT ALDHs and a plant BADH.";
RL Gene 103:45-52(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97426167; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,

```

```

RA Lin D., Namath A., Oefner P., Roberts D., Schraam S., Davis R.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS A SOLUBLE NAD-DEPENDENT BETAINATE ALDEHYDE-
CC SPECIFIC DEHYDROGENASE.
CC -!- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +
CC NADH.
CC -!- PATHWAY: Osmoregulatory choline-glycine betaine pathway.
CC -!- SUBUNIT: Homotrimer.
CC -!- INDUCTION: By osmotic stress. Choline is required for full
CC expression.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52905; CAA37092.1; --
DR EMBL; M77739; AAA23506.1; --
DR EMBL; M77739; AAA23505.1; ALT_INIT.
DR EMBL; AB000138; AAC73415.1; --
DR EMBL; U73857; AAB18038.1; ALT_INIT.
DR PIR; S15181; S15181.
DR HSSP; P56533; 1A4S.
DR EcoGene; EG10110; betB.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldehyd; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Complete proteome.
FT INIT MET 0
FT NP_BIND 228 233 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 251 251 BY SIMILARITY.
FT ACT_SITE 285 285 BY SIMILARITY.
FT CONFLICT 231 231 A -> R (IN REF. 2).
FT CONFLICT 311 311 R -> P (IN REF. 2).
SQ SEQUENCE 489 AA; 52780 MW; E0F7B33220DB2789 CRC64;

Query Match 27.5%; Score 658.5; DB 1; Length 489;
Best Local Similarity 31.9%; Pred. No. 6.3e-33;
Matches 157; Conservative 101; Mismatches 199; Indels 35; Gaps 9;

QY 4 BYQNYNGSEKSSVN-QIEILSPIDDSLGFPVAMTRREVDHAKAGREALPAWALIV 61
Db 5 EQQYIHGYSATSGRTFTINPANGVNLATVQAGREDVDEAVKSAQGGKQKWSMTA 64
QY 62 YERAOYLHKAADIIRDKBEIATVLAKEISKAYN-ASVTEVVTADLIRYAAEGIKLST 120
Db 65 MERSRILRAVDILRERNDLAKLETLDTGKAYSETSVDIVTGADVLEYVA----- 116
QY 121 SADEGGKMDASTGHKLAV-----IRROPVIGVILAIAPNYPVNLSSGIAPALIGNV 173
Db 117 -----GLIPALEGSGPIRSTSVYTRREPLGVAGVAGWNPYQIALWKAPALAAQNA 171
QY 174 VMFKPPTQGSVGLVLAKAFARAGLPAGVFNITGRSGEIGDVIVHEEVNFINTGSTP 233
Db 172 MIFKPEVTPPLTALKLAEIYSEAGLPDGVFNVLGVGAETGQVLTGFCIAKVSFTGVA 231
QY 234 VGORI---GKLAGEPIMLEGGKAGIVLADADLNNAAQIVAGAYDYSGQRTAKRV 290
Db 232 SGKKVWANGSAASLKEVTMELGGKSLPIVFDADLDAADIANMANFFSSQGVCTNTRV 291
QY 291 LVVEVADLAELAKISENVAKLVSQDPFDNAT-VTPVIDDINSADFTESLVVDAROKGAKEL 349
Db 292 FVPACKCAAFEQKILARVERIRAGVDVFDQNFGLVSPFHRDNLVRIAKGKEGARVL 351
QY 350 ---NEFKRD-----GRLLTPGLFDHVTLMKLMWEPFPIPIRIVKDAEAVAIANKSD 402
Db 352 CGGVLVKGGDFNGAWVAPVFTDCSDDDMTIVREILFGVWMSILITVESEDEVIRRANDTD 411

```


QY 472 LDM 474
 Db 505 IKL 507

RESULT 15
 DHAL_ECOLI
 ID DHAL_ECOLI STANDARD; PRT; 495 AA.
 AC P23683; P78250;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 28-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative aldehyde dehydrogenase (EC 1.2.1.3).
 GN ALDH OR B1300.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91216440; PubMed=1840553;
 RA Heim R., Strehler E.F.;
 RT "Cloning an Escherichia coli gene encoding a protein remarkably
 RT similar to mammalian aldehyde dehydrogenases.";
 RL Gene 99:15-23 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9778503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
 RA Kasai H., Kishimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Mocomura K.,
 RA Nakabe S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377 (1996).
 CC -!- FUNCTION: WEAK BUT MEASURABLE ALDH ENZYME ACTIVITY THAT PREFERS
 CC NADP+ OVER NAD+ AS COENZYME.
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -!- PATHWAY: Ethanol utilization; second step.
 CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; M38433; AAC23428.1; -.
 CC ENBL; AE000228; AAC74382.1; -.
 CC ENBL; D90768; BAA14869.1; -.
 CC PIR; G64878; G64878.
 CC HSP; P20000; IAG8.
 CC EcoGene; EG10036; aldh.
 CC InterPro; IPR002086; Aldehyde_dehydr.
 CC Pfam; PF00171; aldedh.1.
 CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

KW Oxidoreductase; NAD; Complete proteome.
 FT NP BIND 244 249 NAD (ADP PART) (BY SIMILARITY).
 FT ACT SITE 267 BY SIMILARITY.
 FT ACT SITE 302 BY SIMILARITY.
 FT CONFLICT 313 S -> R (IN REF. 1).
 SQ SEQUENCE 495 AA; 53418 MW; A20929C55F51C709 CRC64;

Query Match 27.0%; Score 647; DB 1; Length 495;
 Best Local Similarity 33.3%; Pred. No. 3.2e-32;
 Matches 164; Conservative 95; Mismatches 203; Indels 30; Gaps 11;

QY 1 LTKYQNYVNGKSSVQ--LEILSPIDSSLGVPVPMATREEDVHAKAGREALPA--W 56
 Db 16 LAIENRLFINGEYTAANEFTFVDTQAPLAKAGKSVDDIRAMGAAGVFERGDW 75
 QY 57 AALTVERAOYLHKAADIIRDKKEIATVLAKESIKAYNASV-TEVVRADLIRYAAEEG 115
 Db 76 SILSPARKKAVLNKLADLMEAHABELALLETLDTCPIRHSRLDDIPGAARAIKYAE-- 133
 QY 116 IRLTSADG--GGMDASTGKLAIVIRQPVGVIAIAPYNPVNLGSKIAPALIGNVV 174
 Db 134 ----AIDKVGVEVATTSHELAMIVREPVGIVAAIVPWNFPFLLTCKLGPALAGNSV 188
 QY 175 MEKPTGSGVSLVAKAFAGVAGVNTITGRSGEIGDYIVHEEVEVNFNFTGTPV 234
 Db 189 ILKPSKSPUSAIRLAGLAKAGLPDGVNLNVVTFGFGAGQALSRNDIDALFTGSTR 248
 QY 235 GORIKLAG---MRPIMLELGGKDGAGIVLADA-DLDNAAKQIVAGAYDYSGORCTAIRKV 290
 Db 249 GKQLLKDGSDNMRKVLWLEAGKSNIVFADCPDLQQAASATRAAGIFYNQGVCIAGTRL 308
 QY 291 LVVEVDAELAEKISENVAKUSVGDPPDNA-TVTVIDDNGADFIESLVVDARQKAGEL 349
 Db 309 LLEESIADEEFALLKQQAQNWQPGHPLDPATTMTGLIDCAHADSVHSFIREGESG--- 364
 QY 350 NEFKDGR-----LITPLGLFDHVTLMKLAWEPFGPILPIRVKMDAEAVAIANKSDF 403
 Db 365 -QLLDGNAGLAAIGFTIFVDVDPNASLSREEIFGVVLVTFSEEQALQLANDSQY 423
 QY 404 GHOQSVTRDFQAFDIANKLEVGTIVHNNKTRGQPDNFPFLGLKSGAGVQGIYSIEA 463
 Db 424 GLGAAVWTRDLSSAHMSRRLKAGSVFVNNYND-GDMVTPFGYKQSGNGRDKSLHALEK 482
 QY 464 MYNKSVLDMK 475
 Db 483 FTELKTIWISLE 494

RESULT 16
 DHAS_CHICK
 ID DHAS_CHICK STANDARD; PRT; 499 AA.
 AC O93344;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (Retinaldehyde-specific
 DE dehydrogenase type 2) (RALDH(II)) (RALDH-2).
 GN ALDH1A2 OR RALDH2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98394476; PubMed=9727493;
 RA Sockanathan S., Jessell T.M.;
 RT "Motor neuron-derived retinoid signaling specifies the subtype
 RT identity of spinal motor neurons.";
 RL Cell 94:503-514 (1998).
 CC -!- FUNCTION: Recognizes as substrates free retinal and cellular
 CC retinol-binding protein-bound retinal. Does metabolize octanal and
 CC decanal but does not metabolize citral, benzaldehyde, acetaldehyde

CC and propanal efficiently (By similarity).

CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.

CC -!- PATHWAY: Retinoic acid biosynthesis.

CC -!- SUBUNIT: Homotetramer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; AF064253; AAC34299.1; -.

DR HSPSP; P51977; LBXS.

DR InterPro; IPR002086; Aldehyde_dehydr.

DR Pfam; PF00171; aldedh; 1.

DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.

DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

KW Oxidoreductase; NAD.

FT NP_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).

FT ACT_SITE 267 267 POTENTIAL.

FT ACT_SITE 301 301 POTENTIAL.

SQ SEQUENCE 499 AA; 54774 MW; F471F09F3D5645A7 CRC64;

Query Match 26.7%; Score 640; DB 1; Length 499;

Best Local Similarity 33.7%; Pred. No. 8.6e-32; Indels 26; Gaps 9;

Matches 163; Conservative 81; Mismatches 216; Indels 26; Gaps 9;

QY 3 KEYONYNGEKKSVN--QIEILSPIDSSIGFVPNTREEVDHAKAGREAL---PAAW 57

DB 17 KHTKIFINWQNSSESRFPVPNATGEQICEIQEADKVDTDKAVAAALFSLGVSVR 76

QY 58 ALTVVERAQYLKAAIDIERDEKEEATV---LAKSEIKANNASVTEVETADLIRYAE 113

DB 77 RMDASERQLDKLADIVERDRAVLATWESINGKPFLOAFYVDLQGVKT---LRYTAG 133

QY 114 EGRILSTADSGGMDASTGHKLAVIRPQVGIVLAIAPYVPVNLSSGIAPALIGNV 173

DB 134 WADKIH-----GMTIPVDGDTFTTRHEPIGVCGQLIPWNPFLMLFAWKIAPALCCGT 187

QY 174 VMFKPPTOGSVGLVAKAFABAGLPAGVFNITITGRSGEIGDYIVHEEVNFTGTSTP 233

DB 188 VIKPAEQTPLSALYMGALIKKAGPPGCVNLLPGFPIVGAATASHVGIDKIAFTGST 247

QY 234 VQQRIGKLAG---MRPIMLELGGKAGIVLADADLIDNAKQIVAGAYDSGQECTAIRV 290

DB 248 VCKLIQEAAGRSNLRKVTLELGGKSPNIIIFADADLDYAVEQAHQGVFNQGQCCTAGSRI 307

QY 291 LVVEEVADELAEKISENVAKLVSQDPPDNATVT-PVIDDNSADFTESLVVDARQKAKEL 349

DB 308 YVEESIEFVRVSVERARRVVGSPDPTTEQGPDLKKQYNKILIELSGITGAKLE 367

QY 350 NEFK---RDGRLTLPGLDRVLTDMKLAWEEPFGPIPIIRVQAEBAVAIANKSDFLQ 406

DB 368 CGKGLGKRGKFFIPTVFSNVDDMRIAKEEITFGPQVEILRFKTVDEVIERNANSDFGLV 427

QY 407 SVVFTDRQKAFDIANKLVGVTHVNNKTRGPDNPPFLGLKSGAGVQGIIRYSIEMTN 466

DB 428 AAVFTNDINKALTVSMAQGTWVINCYNALNAQS-PFGGFMKSGNGREMGSLREYSE 486

QY 467 VKSIVL 472

DB 487 VKTVII 492

RESULT 17

DNA2 HUMAN

ID DNA2 HUMAN STANDARD; PRT; 499 AA.

AC O94786;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (Retinaldehyde-specific

DE dehydrogenase type 2) (RALDH2) (RALDH-2).

GN ALDH1A2 OR RALDH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99038200; PubMed=9819382;

RA Ono Y., Fukuhara N., Yoshie O.;

RT "TALI and LIM-only proteins synergistically induce retinaldehyde

RT dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by

RT acting as cofactors for GATA3.";

RL Mol. Cell. Biol. 18:6939-6950(1998).

CC -!- FUNCTION: Recognizes as substrates free retinal and cellular

CC retinol-binding protein-bound retinal. Does metabolize octanal and

CC decanal but does not metabolize citral, benzaldehyde, acetaldehyde

CC and propanal efficiently (By similarity).

CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.

CC -!- PATHWAY: Retinoic acid biosynthesis.

CC -!- SUBUNIT: Homotetramer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; AB015226; BAA34785.1; ALT_INIT.

DR HSPSP; P51977; LBXS.

DR Genew; HGNC:15472; ALDH1A2.

DR MIM; 603687; -.

DR GO; GO:0004028; F:aldehyde dehydrogenase activity; TAS.

DR GO; GO:0003824; F:catalytic activity; TAS.

DR GO; GO:0005489; F:electron transporter activity; TAS.

DR GO; GO:0006776; P:vitamin A metabolism; TAS.

DR InterPro; IPR002086; Aldehyde_dehydr.

DR Pfam; PF00171; aldedh; 1.

DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.

DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

KW Oxidoreductase; NAD.

FT NP_BIND 244 249 NAD (ADP PART) (BY SIMILARITY).

FT ACT_SITE 267 267 POTENTIAL.

FT ACT_SITE 301 301 POTENTIAL.

SQ SEQUENCE 499 AA; 54780 MW; FE0B53A47644246B CRC64;

Query Match 26.7%; Score 639; DB 1; Length 499;

Best Local Similarity 33.7%; Pred. No. 9.8e-32; Indels 26; Gaps 9;

Matches 162; Conservative 81; Mismatches 212; Indels 26; Gaps 9;

QY 8 YVNGEKKSVN--QIEILSPIDSSIGFVPNTREEVDHAKAGREAL---PAAWALTVY 62

DB 22 FINEWQNSSESRFPVPNATGEQICEQVQADKADIDKAVQAALFSLGVSWRMDAS 81

QY 63 ERAQYLKAAIDIERDEKEEATVLA---KEISKANNASVTEVETADLIRYAEIGRL 118

DB 82 ERGRLLDKLADIVERDRAVLATWESINGKPFLOAFYVDLQGVKT---FRYAGWADKI 138

QY 119 STNDSGGGMDASTGHKLAVIRPQVGIVLAIAPYVPVNLSSGIAPALIGNVNVEKP 178

DB 139 H-----GMTIPVDGDTFTTRHEPIGVCGQLIPWNPFLMLFAWKIAPALCCGTWVIRP 192

QY 179 PTGSGVSGVLAKAFABAGLPAGVFNITITGRSGEIGDYIVHEEVNFTGTSTPVGQRI 238

DB 193 AEQTPLSALYMGALIKKAGPPGCVNLLPGFPIVGAATASHIGIDKIAFTGTSTGVKLI 252


```

RN  SEQUENCE.
RP  TISSUE=Liver;
RC  MEDLINE=98034175; PubMed=9369242;
RX  Hjeltnvist L., Lundgren R., Norin A., Joernvall H., Vallee B.,
RA  Klyosov A., Keung W.M.;
RT  "Class 2 aldehyde dehydrogenase. Characterization of the hamster
RT  enzyme, sensitive to daidzin and conserved within the family of
RT  multiple forms.";
RL  FEBS Lett. 416:99-102(1997).
CC  -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC  -|- PATHWAY: Ethanol utilization; second step.
CC  -|- SUBUNIT: Homotetramer.
CC  -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC  -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
DR  HSP; P05091; 1CW3.
DR  InterPro; IPR002086; Aldehyde_dehydr.
DR  Pfam; PF00171; aldedh; 1.
DR  PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR  PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
KW  Oxidoreductase; NAD; Mitochondrion.
FT  NP BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT  ACT_SITE 268 268 BY SIMILARITY.
FT  ACT_SITE 302 302 BY SIMILARITY.
SQ  SEQUENCE 500 AA; 54334 MW; 8B8186AB04F493F CRC64;

Query Match 26.3%; Score 630; DB 1; Length 500;
Best Local Similarity 34.1%; Pred. No. 3.5e-31;
Matches 164; Conservative 83; Mismatches 212; Indels 22; Gaps 9;

QY 6 QNYVNGEWSVNO--TEILSPDSSGLGFPVPMTRREVDHAKAGREAL---PAAALIT 60
Db 21 QIFNNEWDHDAVSKTFTTNPSTGEVQVAGSKEDVDKAVKAAAFOLGSPRRMD 80
QY 61 VYERAOVLHKAADIIRKKEIATVLAKESIKAYNAS-VTEVVRADLIIRYABEGIRLS 119
Db 81 ASDRGLLNRADLIIRDRYLALETLDNGKPVISYVLDLMDVLKCRYYA----- 133
QY 120 TSADE-GGRDASTGCHKLAVIRQPVGIVLAIFAPYPNVNLSSKIAPALIGNVVFXP 178
Db 134 GWADKYHGKTIPIDGPFSYTRHEPVGCGQIIPWPFLLMQAKLGPALATGNVVMKV 193
QY 179 PTQSGVGLVLAFAEAGLPAGVFNITGRSGEIGYIVHEHVNFINFTGSTPGQRI 238
Db 194 AEQTPLTALYVANLIKANGPPGPGVNVNVEGFGTAGAAJASHEDVDKVAFTGSTEVGHLI 253
QY 239 GKLAG---MRPIMLELGGKAGATVLADADLNAKQIVAGDYSGQCTAIKRVLVVEE 295
Db 254 QVANGSNLKRVTLELGGKSPNIIIMSDADMWAVEQAHFALFNGQCCAGSRTFVED 313
QY 296 VADELAEKISENVAKLSVGGPFDNATVT--PVIDNSADFTESLVDVDAQKAKEL---NE 351
Db 314 VYDEFVRSVARAKSRVGNPFDSRTSQPQVDETFQFKILGVIKSGQEGAKLLCGGA 373
QY 352 FKRDGRLTPGLFDHVIDMLKAWPEFCPLPIIRYKDAEAVAIANKSPFGLSQSVT 411
Db 374 ADRGVTIQTPTVGGVXDGMTIAKEEIEFGVWQILKFTIEEVGRANNSKYGLAAAVT 433
QY 412 RDFQKAFDIANKLEVGVTHNNKTGRGPNPFPLGLKSGAGVGIRYSTEAMTNVKSIV 471
Db 434 KOLDKANYLSQALQNGTWINCYDVEGAQS-PFGYKXSGSGRELGEYGLQVTEVKTVT 492
QY 472 L 472
Db 493 I 493

RESULT 24
ID DHAM RAT
AC P11884;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH
DE class 2) (ALDH1) (ALDH-E2).
GN ALDH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=89210865; PubMed=2540003;
RA Farres J., Guan K.-L., Weiner H.;
RT "Primary structures of rat and bovine liver mitochondrial aldehyde
RT dehydrogenases deduced from cDNA sequences.";
RL Eur. J. Biochem. 180:67-74(1989).
RN [2]
RP SEQUENCE OF 1-29 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88134217; PubMed=3342060;
RA Farres J., Guan K.-L., Weiner H.;
RT "Sequence of the signal peptide for rat liver mitochondrial aldehyde
RT dehydrogenase.";
RL Biochem. Biophys. Res. Commun. 150:1083-1087(1988).
RN [3]
RP SEQUENCE OF 1-19.
RC TISSUE=Liver;
RX MEDLINE=91378548; PubMed=1898068;
RA Jung J., Weiner H.;
RT "Purification and characterization of catalytically active precursor
RT of rat liver mitochondrial aldehyde dehydrogenase expressed in
RT Escherichia coli.";
RL Arch. Biochem. Biophys. 289:214-222(1991).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -|- PATHWAY: Ethanol utilization; second step.
CC -|- SUBUNIT: Homotetramer.
CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL; X14977; GAA33101.1; -.
DR EMBL; M19030; AAA40719.1; -.
DR PIR; S03564; S03564.
DR PDB; 1OM2; 15-MAR-00.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation; Mitochondrion; Transit peptide;
KW 3D-structure.
FT TRANSIT 1 19 MITOCHONDRION.
FT CHAIN 20 519 ALDEHYDE_DEHYDROGENASE.
FT MOD_RES 20 20 ACETYLATION (PROBABLE).
FT NP_BIND 264 269 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 287 287 BY SIMILARITY.
FT ACT_SITE 321 321 BY SIMILARITY.
SQ SEQUENCE 519 AA; 56488 MW; 75C74820F133B5 CRC64;

Query Match 26.1%; Score 625; DB 1; Length 519;
Best Local Similarity 33.9%; Pred. No. 7.3e-31;
Matches 163; Conservative 83; Mismatches 213; Indels 22; Gaps 9;

QY 6 QNYVNGEWSVNO--TEILSPDSSGLGFPVPMTRREVDHAKAGREAL---PAAALIT 60
Db 40 QIFNNEWDHDAVSKTFTTNPSTGEVQVAGSKEDVDKAVKAAAFOLGSPRRMD 99

```

```

SQ SEQUENCE 500 AA; 54166 MW; 7F95364A73383B24 CRC64;

Query Match          26.1%; Score 624; DB 1; Length 500;
Best Local Similarity 33.7%; Pred. No. Ge-31;
Matches 162; Conservative 85; Mismatches 212; Indels 22; Gaps 9;
```

Qy	6	QNYVGEWKSVMNQ--IBILSPIDDSGLGFPVPMTRTEEVDDHAKAGREAL---	PAAALT 60
Db	21	QIFINNEHWADVASKKTFTVTNVESTGEVICQVAAGDKEDVDRAKARAFAAQFGSGPWRMD 80	:
Qy	61	VYERAOYLHKAAADI IERKEEIAIVLAKESIKAYNAS-VTEVVVRTADLIIRYAEEGIRLS 119	:
Db	81	ASDRGLLNRLIADIERDRTYLAALETLDNGKPVIISYLVLDMLVKCLRYA----- 133	:
Qy	120	TSADE-GGKMADSTGHKLAVIRROPVGIVLATAPNPVPNLGSKIAPALICGNVMPKP 178	:
Db	134	GWADKYHGKITPIDGFFSYTRHEPVGVCQIIIPNPFLLMQAAKLGPA LATGNVWMKV 193	:
Qy	179	PTQGSVSLGVLA KAFAEAGLPAGVNTITGRGSEIGDIYIVEHEVNFINPTGTSPVGORI 238	:
Db	194	AEQTLTALTVALTKAEAGFPFGPVNVVPGFGPTAGAASHEDVDKVAFTGSTVEGHLI 253	:
Qy	239	GKLAG--MRPIMLEGGKGDAgilADADOLDNAAKOIVAGADYSGORCTAIKRVLVWEE 295	:
Db	254	QVAAGRSLKKVLTLELGGGSNNIIVSDAMDWAQEHAFLFNQGCGCAGSRFTVQED 313	:
Qy	296	VADELAEKISENVAKLSVGDPDFNATVT--PVIDDNSADFTESLVDVAROKAGEL---NE 351	:
Db	314	VVAEPVERSARAKSRVGVNFSDQSQTQGPQVDETQFNKVLGYIKSGKEGAKLLCGGA 373	:
Qy	352	FYRDGRLITPLPDHVTLLDMKLAWEPFGPIPIPIRVKORAEVAIAINKSDPGLSSVFT 411	:
Db	374	AADRGYFTQTFEGVDQGMITAKESIFGPVWLKFKEITEEVGRANNSKYGLAAAVFT 433	:
Qy	412	RDFQKAPDIANKLEVGTTHINKTGDPNPNFPLGLKSGAGVGQIRYSTTEAMTVKSVI 471	:
Db	434	KOLDKANVLSQALOGXTWINCTDVFGAGS-PFGYGVMSGNRELGEYGLOATEVTKVT 492	:
Qy	472	L 472	:
Db	493	I 493	:

RESULT 26	DHAI RAT	STANDARD;	PRT;	500 AA.
ID	DHAI RAT	STANDARD;	PRT;	500 AA.
AC	P51647;	O09184;		
DT	01-OCT-1996	(Rel. 34, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	15-MAR-2004	(Rel. 43, Last annotation update)		
DE	Aldehyde dehydrogenase IA1 (EC 1.2.1.3)	(Aldehyde dehydrogenase,		
DE	cytosolic)	(ALDH class 1) (Retinal dehydrogenase) (RALDH) (RALDH1).		
GN	ALDH1A1 OR ALDH.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=Sprague-Dawley; TISSUE=Kidney;			
RC	MEDLINE=96125208; PubMed=854180;			
RC	Bhat P.V., Labrecque J., Boutin J.-M., Lacroix A., Yoshida A.;			
RT	"Cloning of a cDNA encoding rat aldehyde dehydrogenase with high			
RT	activity for retinal oxidation.";			
RL	Gene 166:303-306 (1995).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=Sprague-Dawley; TISSUE=Liver;			
RC	MEDLINE=97212790; PubMed=9059608;			
RC	Kachmann E.C., Lipsky J.J.;			
RA	"A preliminary report on the cloning of a constitutively expressed rat			
RA	liver cytosolic ALDH cDNA by PCR.";			
RL	Adv. Exp. Med Biol 414:69-72(1997).			

or send an email to license@sib-sib.ch).

CC
EMBL; M74570; AAA37202.1; -
DR EMBL; M74571; AAA37203.1; -
DE EMBL; S75713; AAB32754.2; -
OS EMBL; S77047; -; NOT_ANNOTATED_CDS.
OC PIR; JQ1004; JQ1004.
OC HSP; F51977; IBSX.
RN SWISS-2DPAGE; P24549; MOUSE.
DR MGD; MG1.1353450; Aldh1al.
GO GO:G0004028; F:aldehyde dehydrogenase activity; IDA.
GO GO:G0042573; P:retinoic acid metabolism; IMP.
InterPro: IPR002086; Aldehyde_dehydr.
Pfam; PF00171; aldehyd; 1.
PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation.

FT INIT MET 0 BY SIMILARITY

FT MOD RES 1 ACETYLATION (BY SIMILARITY)

FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY)

FT ACT_SITE 268 268 BY SIMILARITY

FT ACT_SITE 302 302 BY SIMILARITY

FT BINDING 299 304 ANTABUSE

FT CONFLICT 7 7 A -> R (IN REF. 1)

FT CONFLICT 44 44 T -> S (IN REF. 2)

FT CONFLICT 50 50 H -> Q (IN REF. 2)

FT CONFLICT 86 86 R -> C (IN REF. 1)

FT CONFLICT 457 457 I -> M (IN REF. 2)

SQ SEQUENCE 500 AA; 54318 MW; 3E428154E7214B54 CRC64;

Query Match 26.0%; Score 622; DB 1; Length 500;
Best Local Similarity 32.8%; Pred. No. 1.1e-30;
Matches 156; Conservative 92; Mismatches 210; Indels 20; Gaps 8;

QY 8 YVNGEKSVN--QIEILPSDDSLGFVPMTREVHAKMGREAL---PAWAALTVY 62
DB :|::||::||::||::||::||::||::||::||::||::||:
DQ FINNEHNVSCKKFPVLNPATEVI CHEBGKDADVDAVKAAQAPQSPTWDAS 82
DY :|::||::||::||::||::||::||::||::||::||:
QY 63 ERAQLHKADIIDKEELATFLAKEISKAY-NASVTVEVRTADLRVAEEGRLSTS 121
DB ||:||::||::||::||::||::||::||::||::||::||:
DQ ERGLINKLADLMERDLALTMEALNGKVFNANVALSDLGGCIKALKYCAGWADKH-- 140
DY ::||::||::||::||::||::||::||::||::||::||:
QY 122 ABEGKMDASTGHKLAVTRQPVGIVAIAPYNPNVSGSKTAPALIGNVNFKPTQT 181
DB :|::||::||::||::||::||::||::||::||::||:
DQ 141 ----GQTIPSDGDIFFYTTEREPGVCGQIIPNNFPLMFIWKIGPALSCGNVVKKPAEQ 196
DY :|::||::||::||::||::||::||::||::||::||:
QY 182 GSVSLGLAKAFAGAEGLPAGVENITTCRGSELDGYIVEHEVNFINFTGTFVQRIGKL 241
DB :|::||::||::||::||::||::||::||::||::||:
QY 197 TLPTALTHLASLIKEAGFGPGVUNVPOVGTGAGALSSEMVDVKFAFTSQVQKLKA 256
DB :|::||::||::||::||::||::||::||::||::||:
QY 242 AG---MRPIMLLGKGADGIVLADADLNAAQIVAGAYDYSQRTALKRVLVVEEAD 298
DB :|::||::||::||::||::||::||::||::||::||:
QY 257 ACKSNLKRVTLLEGKSICVFADADDLAJAVEFAHHGVFYHQGCCVAARSIFVESYVD 316
DB :|::||::||::||::||::||::||::||::||::||:
QY 299 ELAEKISENAKLSVCDFP-DNAVTPVIDDNSADFTESLVDARKGAK---ELENFKR 354
DB :|::||::||::||::||::||::||::||::||::||:
QY 317 EFVKSESVERAKTYVLGNPTLPINQGPDQDKEDHLIDLIESGKCEAKLECGGRWG 376
DB :|::||::||::||::||::||::||::||::||::||:
QY 355 DGRLLTPLGFDHTVLTDMKIAEEPFGILLPIIRVVDKAEBAVANANKSDFGLSSVTRDF 414
DB :|::||::||::||::||::||::||::||::||::||:
QY 377 KGFFVOFVNSVNTDEMRAKEEIIFGVQQIMKFSVDVVKRANNITYGAAGLGFTKOL 436
DB :|::||::||::||::||::||::||::||::||::||:
QY 415 OKAFDIANKLEVSTGHNNKTGRGPDFNPFPLGLKSGAGVQGIKYRSYEAMTNVKSIVL 472
DB :|::||::||::||::||::||::||::||::||::||:
QY 437 DKAITVSSALQAGVWNVCYMLSA-QCPFGGFKMSGNGRELGEHCLVEYTELKTVM 493
DB :|::||::||::||::||::||::||::||::||::||:

RESULT 28
DHAM_MOUSE
AC P47738;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

[illegible]

[9] SEQUENCE OF 214-500 FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=87279033; PubMed=3610592;
 RA Agarwal D.P., Goede H.W.,
 RT "Human aldehyde dehydrogenase isozymes and alcohol sensitivity.";
 RL Isozymes Curr. Top. Biol. Med. Res. 16:21-48(1987).
 RN [10]
 RP DESCRIPTION OF ORIGIN OF CONFLICTS BETWEEN REF.2 AND DNA SEQUENCES.
 RP MEDLINE=88005159; PubMed=3653404;
 RA Hempel J., Hoeeg J.-O., Joernvall H.;
 RT "Mitochondrial aldehyde dehydrogenase. Homology of putative targeting
 RT sequence to that of carbamyl phosphate synthetase I revealed by
 RT correlation of cDNA and protein data.";
 RL FEBS Lett. 222:95-98(1987).
 RN [11]
 RP VARIANT LYS-504.
 RP MEDLINE=84119449; PubMed=6582480;
 RA Yoshida A., Huang I.-Y., Ikawa M.;
 RT "Molecular abnormality of an inactive aldehyde dehydrogenase variant
 RT commonly found in Orientals.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:258-261(1984).
 RN [12]
 RP VARIANT LYS-496.
 RP MEDLINE=96119362; PubMed=8561277;
 RA Novorodovsky A., Tsai S.J., Goldfarb L., Peterson R., Long J.C.,
 RA Goldman D.;
 RT "Mitochondrial aldehyde dehydrogenase polymorphism in Asian and
 RT American Indian populations: detection of new ALDH2 alleles.";
 RL Alcohol. Clin. Exp. Res. 19:1105-1110(1995).
 RN [13]
 RP X-RAY CRYSTALLOGRAPHY (2.58 ANGSTROMS).
 RA MEDLINE=20095857; PubMed=10631996;
 RA Ni L., Zhou J., Hurley T.D., Weiner H.;
 RT "Human liver mitochondrial aldehyde dehydrogenase: three-dimensional
 RT structure and the restoration of solubility and activity of chimeric
 RT forms.";
 RL Protein Sci. 8:2784-2790(1999).
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H2O = an acid + NADH.
 CC -!- PATHWAY: Ethanol utilization; second step.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- POLYMORPHISM: Allele ALDH2*2 is associated with a very high
 CC incidence of acute alcohol intoxication in Orientals and South
 CC American Indians, as compared to Caucasians.
 CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC -!- CAUTION: The conflicts between the sequence determined in Ref.1
 CC and Ref.2 and that in Ref.3 and Ref.4 are probably all due to
 CC frameshift or sequencing errors as described in Ref.5 and Ref.6;
 CC the sequence described in Ref.9 differs from that shown due to two
 CC frameshifts.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/ALDH2ID250.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: M20454; AAA51693.1; JOINED.
 DR EMBL: X05409; CAA28990.1; -.
 DR EMBL: Y00109; CAA68290.1; -.
 DR EMBL: BC002967; AAH02967.1; -.
 DR EMBL: M03001; AAB59500.1; -.
 DR EMBL: M26760; AAA51694.1; -.
 DR EMBL: M54931; AAA62825.1; ALT_FRAME.
 DR PIR: A29975; DEHUE2.
 DR FDB: ICW3; 26-SEP-01.
 DR FDB: I005; 04-MAR-03.
 DR Genew; HGNC:404; ALDH2.
 DR CK; P05091; -.
 DR MIM; 100650; -.
 DR GO; GO:0004029; F:aldehyde dehydrogenase (NAD) activity; TAS.
 DR GO; GO:0004030; F:aldehyde dehydrogenase (NAD(P)) activity; TAS.
 DR GO; GO:0005489; F:electron transporter activity; TAS.
 DR GO; GO:0006066; P:alcohol metabolism; TAS.
 DR GO; GO:0005975; P:carbohydrate metabolism; TAS.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR PROSITE; PS00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism;
 KW 3D-structure.
 FT TRANSIT 1 17 MITOCHONDRION.
 FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE.
 FT NP_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 285 285
 FT ACT_SITE 319 319
 FT VARIANT 337 337
 FT VARIANT 496 496 E -> V (in dbSNP:1062136).
 FT VARIANT 504 504 E -> K (in allele ALDH2*3).
 FT VARIANT 504 504 E -> K (in allele ALDH2*2; drastic
 FT reduction of enzyme activity; dbSNP:671).
 FT CONFLICT 7 12 RFGRL -> ARAPP (IN REF. 5).
 FT CONFLICT 7 26 RGERGLRLSAAATQAVP -> AWPAAWPLVSRHPCR
 FT CONFLICT 18 18 A (IN REF. 3).
 FT CONFLICT 80 85 VKAARA -> REGRPG (IN REF. 3 AND 4).
 Query Match 26.0%; Score 621; DB 1; Length 517;
 Best Local Similarity 33.7%; Pred. No. 1,3e-30;
 Matches 162; Conservative 85; Mismatches 212; Indels 22; Gaps 10;
 QY 6 QNTVNGEKKSSVNO--LEILSPIDSSLGFPVPMTRVEDVHAKAGREAL---PAWAALT 60
 Db 38 QIFINNEHDAVSRKTPPTVNPSTGEVICQVAEGDKEDVDKAKAARAAAFQLGSPWRMD 97
 QY 61 VYBPAQYLKKAADIIRKKEETATVLAKETSKAYNAS-VTEVVRTADLIRYAAEGIRLS 119
 Db 98 ASHREGLNLRADLIEDRTYLAETUDNGKPVISYLDVLDWVLCILRYA----- 150
 QY 120 TSADE-GGKDMDASTGHKLAVIRROPVGIYLAIAIPYVNLSSGKIAPALITGNVWFEP 178
 Db 151 GWADKYHGKTIPIIDGDFSVTRHEPVGVCGQIIPWFLPQAKLGPALATGNVWVKV 210
 QY 179 PTQGSVGLVIKAFAPAGIPAGVNTITGRGSEIGDYIVEHEVNFINTGTPVGQRI 238
 Db 211 AEQTPLTALYANLILKEAGFPFPGVNVIVPGFTAGAAIASHEDVDKVAFTGSTEIGVI 270
 QY 239 GKLAG---MRPIMLELGGKDAGIVLADADLDNAKQIVAGADYSQRCCTAIKRVLVVEE 295
 Db 271 QVAASSNLKRVLTLEGGKSPNIMSDADMWAVEQAHFALFFNQGCCAGSRTFQED 330
 QY 296 VADELAARKISENVAKLSVGDPDFDNATVT--FVIDDMSADFIESLVVDARQKAKEL--NEF 352
 Db 331 IYDEFVRSVARAKSRVVGPNPDSKTSQGFQVDETFQKKILGYINTKGQAKLCCGGI 390
 QY 353 KED-GRLLTLPGLFDHVLDMKLAMEEPFGILPIIRVKDAEEAVAIANKSDPGLSSVFT 411
 Db 391 AADRGYFIQPTVFQDQGMTAKEEIFGPMQLKFKTIEEVVGRANNSTYGLAAAVFT 450


```

DE (BADH).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Amaranthaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=cv. Savoy Hybrid 612;
RX MEDLINE=90207274; PubMed=2320587;
RA Weretilnyk E.A., Hanson A.D.;
RT "Molecular cloning of a plant betaine-aldehyde dehydrogenase, an
RT enzyme implicated in adaptation to salinity and drought.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2745-2749(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Shu W., Ai W., Chen S.;
RL Submitted (PEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +
CC NADH.
CC -!- PATHWAY: Betaine biosynthesis; last step.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M31480; AAA34025.1; -.
CC EMBL; U69142; BAB41696.1; -.
CC PIR; A35994; A35994.
CC PIR; T51173; T51173.
CC HSSP; P05091; 1CW3.
CC InterPro; IPR002086; Aldehyde_dehydr.
CC Pfam; PF00171; aldehyd; 1.
CC NP_BIND; 235 240 NAD (ADP PART) (BY SIMILARITY).
CC ACT_SITE; 257 257 BY SIMILARITY.
CC ACT_SITE; 291 291 BY SIMILARITY.
CC CONFLICT; 424 424 S->F (IN REF. 2).
CC SEQUENCE; 497 AA; 55088240B635B22F CRC64;
CC
CC Query Match 25.8%; Score 617.5; DB 1; Length 497;
CC Best Local Similarity 30.6%; Pred. No. 2e-30;
CC Matches 151; Conservative 98; Mismatches 204; Indels 41; Gaps 9;
CC
CC 6 QNYNGEKSSV--NQIEILSPIDSSLGFPVPMTRVEDVHAKMGREALPA--WAAALTV 61
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 9 QLFIDGWEPIKKNRIPVNPSTETIIGDIPAAEDVEVAVAAARFRNNWSATSG 68
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 62 YERAOYLKKAADIIEERDEETATVLAKEISKAYNASVTEVVRADLIRYAAEEGIRLTS 121
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 69 AHRATYLRALAKITEKHDFVKLETIDSGKPFDEAVLIDDDVASCFFYFA----- 119
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 122 ADEGKMDASTGKHLAVIR-----RQPGVILAIAPYNPVNLGSKIAPALIG 170
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 120 ----GQAEALDGKQKAPVTLPMERFKSHVLRQPLGVGLISPNWYFLLMATWKIAPALAA 175
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 171 GNVVMFKPTQGSVGLVLAKAFABAGLPAQVFNITGRGSEIGDYIVVEHEVNFNFTG 230
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 176 GCTAVLKPSELASVTCLFEGFECVNEVGLPPGVNLITGLGPDAGAPLVSHDPVDKIAFTG 235
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 231 STPVGRIGKLAG--MRPIMELGGKDGAGIVLADADLDNAKQIVAGAYDYSGRCTATK 289

```

```

Db 236 SSATGSKVMAAAQLVKPVTELGKSPFIWVEDYDIDKVVVEITFCFWTNGQICSATS 295
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 289 RVLVVEEVADELAKEISENVAKLSVGDPF--DNATVTPVIDDNGSADFTESLVVDARQKAG 347
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 296 RLLVHESIAAEFVDKLVKWKTKIKISDPFEGRCLGPVISKGQYDKIMKFISTAKSEAT 355
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 348 EL-----NEFKRDRGLLTPGLFDHVTLDKMLAWBPPFGPILPIIRVKDAEAAVANKSD 402
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 356 ILYGSRPEHLKKGYYIPTVITDISTSQIKWEKVEFVGLCVKVTSSSEDAIALANDTE 415
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 403 FGLQSSVTFDFOKAFDANKLEVGTTHINKTGRG--DNFPFLGLKSGAGVCGIRYS 460
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 416 YGLAAAFVSNLDERCERITKALEVGAVMVNCSQ---PCFVQAPMGSGIKSGRGELGSEWG 472
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 461 IEAMTNVKSIVLDM 474
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 473 IQYINLIKQVTQDI 486
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
RESULT 34
DHAL ASPNG STANDARD; PRT; 497 AA.
AC P41751;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
GN ALdehyde dehydrogenase (EC 1.2.1.3) (ALDDH).
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Euryotales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90108706; PubMed=2606357;
RA O'Connell M.J., Kelly J.M.;
RT "Physical characterization of the aldehyde-dehydrogenase-encoding
RT gene of Aspergillus niger.";
RL Gene 84:173-180(1989).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M32351; AAA87596.1; -.
CC HSSP; P05091; 1CW3.
CC InterPro; IPR002086; Aldehyde_dehydr.
CC Pfam; PF00171; aldehyd; 1.
CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC Oxidoreductase; NAD.
CC NP_BIND; 242 247 NAD (ADP PART) (BY SIMILARITY).
CC ACT_SITE; 265 265 BY SIMILARITY.
CC ACT_SITE; 299 299 BY SIMILARITY.
CC SEQUENCE; 497 AA; 53809 MW; 47CA353FFDBA50E8 CRC64;
CC
CC Query Match 25.8%; Score 617.5; DB 1; Length 497;
CC Best Local Similarity 32.2%; Pred. No. 2e-30;
CC Matches 156; Conservative 84; Mismatches 204; Indels 41; Gaps 9;
CC
CC 8 YNGWEKSSV--NQIEILSPIDSSLGFPVPMTRVEDVHAKMGREALP--WAAALTVYER 64
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 23 FIDGEPVKGAEKGTETINPSNEKPIVAVHEATEKDVTAAAKAFEGSWROVTPSTR 82
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 65 AQYLHKKAADIIEERDEETATVLAKEISKAYNASVTEVVRADLIRYAAEEGIRLTSAD 124

```

```

Db 83 GRMLTKLADLFDABILASIEALDNGKSTMAHGDIAGAAGCLRY-----Y 129
QY 125 GCKMDASTG-----HKLAVTRQCPVGIIVLATAPYVNVNLSKTIAPALICGNVWMEK 177
Db 130 GGWADKIHQTIDTNSSETLNYTRHEPIGVCGQIIPWNFLLMNAWKIGPAITAGTNTVVIK 189
QY 178 PPTQSGVGLAKAFABEAGLPAQVNTITGRGSEIGDYIVHEBEVNFNFTGSPVQOR 237
Db 190 TRAQTPLSGLYAANVKEAGIPAGVNVVSGFGRVAGSAISHMDIDKVAFTGSLVGR 249
QY 238 IKLA---GWRPIMLELGGKAGIVLADALDINAQIVVAGAYDSQRCRTAKRVLVE 294
Db 250 ILQAAKSNLKKVTLGLGKSNIVFNDADIDNAISWANFGIFYNHGCCAGSILVOE 309
QY 295 EVADELAELKISENVAKLSVGDPEFNATVT--PVIDDMSADFTESLVVDARQKAKELNEK 353
Db 310 GIYDFIARLKERALQNKQDPPAKDFTQGPQVSQLOFDRIMEVYIQKGDAGATVAVGE 369
QY 354 R---DRLTLPLGLFDHVTLDMLKAWBEPFGPIIPRIKVDAAEAVATANKSDFLQSSVF 410
Db 370 RHGTGEGYFIQFTVTDVTSMDKINQEEIFGVPVTVQKFDVEDAIKIGNSTSVGLAAGIH 429
QY 411 TRDFOKARDIANKLEVTGTHINKTGRGPNF-----PFLGLKSGAGVQGRYRIEAMT 465
Db 430 TKDVTTAIRVSNALRAGTVWNSY-----NLIQVQVPPGFGKESGIGRELGSYALENT 483
QY 466 NVKSI 470
Db 484 QIKAV 488

RESULT 35
DHAC RAT
ID DHAC RAT STANDARD; PRT; 500 AA.
AC P13601;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, cytosolic 1 (EC 1.2.1.3) (ALDH class 1)
DE (ALHDI) (ALDH-EI).
GN ALDH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89327272; PubMed=2753900;
RA Dunn T.J., Koleske A.J., Lindahl R., Pitot H.C.;
RT "Phenobarbital-inducible aldehyde dehydrogenase in the rat. cDNA
sequence and regulation of the mRNA by phenobarbital in responsive
rats.";
RL J. Biol. Chem. 264:13057-13065 (1989).
CC -!- FUNCTION: Binds free retinal and cellular retinol-binding protein-
bound retinal. Can convert/oxidize retinaldehyde to retinoic acid
(BY similarity).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- INDUCTION: By phenobarbital.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-eib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M23995; AAA40718.1; -.

```

```

DR PIR; A32616; A32616.
DR HSSP; P51977; 1BXS.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldehyd; 1
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT INIT MET 0 0 BY SIMILARITY.
FT NP BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
SQ SEQUENCE 500 AA; 54428 MW; D6806A1AE29A0D56 CRC64;

Query Match 25.8%; Score 617; DB 1; Length 500;
Best Local Similarity 32.0%; Pred. No. 2.1e-30;
Matches 153; Conservative 95; Mismatches 210; Indels 20; Gaps 8;

QY 8 YVNGEWSKSVK--QIEILSPIDSSIGFVPAMTTEVEVDHAKAGREAL---PAWALTY 62
Db 23 FINNEHNSLNGKFFVINPATEEIVCHEBGDKADVDKAVKAAQAFQIGSPWRTMDAS 82
QY 63 BRAQLHKAADIIERDKEEATVLAKEISKAY--NASVTEVVRTADLIRYAAEGIRLST 121
Db 83 ERCLINKLADLMDERDVLATWESNNAGKIFTHAYLLDTEVSIKALYFAGWADKIH-- 140
QY 122 ADEGGKMDASTGHKLAVIRPQVIGVLAIPYVNVNLSKTIAPALICGNVWVPKPTQ 181
Db 141 ---GQTIPSDGDVFTYTRREPVGCGQIIPWNGFLILFIWKIGALISCGNTVIVKPAQ 196
QY 182 GSVGLVLAKAFAPAGLPAQVNTITGRGSEIGDYIVHEBEVNFNFTGSPVQORIGKL 241
Db 197 TPLTALYNASLIKEAGFPFGVNVNVPFGYSTAGAAISSHMDIDKVSFTSGTEVKLIKEA 256
QY 242 AG---MRPTMLELGGKDGIVLADADLNAAKQIVAGAYDSGQRCCTAKRVLVVEVAD 298
Db 257 AGSNLKRVTLELGGKSPCLVFDADLDSAVEFAHQVFFHQGQICVAASELFEESIYD 316
QY 299 ELAEKISENVAKLSVGDPEFNATVT--PVIDDMSADFTESLVVDARQKAGK---ELNEFKR 354
Db 317 EFVRSVERAKKVLGNFLDSDGISQGPQIDKEQHAQKILDLIESGKKGAKLECGGRWGN 376
QY 355 DGRLLTPGLFDHVTLDMLKAWBEPFGPIIPRIKVDAAEAVAIANKSDFLQSSVTRDF 414
Db 377 KGFVQPTVFNVDENRIAKSEIFGPVQIMFKSIDIVIKRANNTFYGIAAGVFTKDL 436
QY 415 QKAFDIANKLVGTWTHNNKTGRGPNPEPLGLKSGAGVQGRYRIEAMTVKSVL 472
Db 437 DRAITVSSALQAGTVWNCVILSV-QCPFGGFNMGNGREMGEGQVYETELKTVAM 493

RESULT 36
DHAC ENCBU
ID DHAC ENCBU STANDARD; PRT; 497 AA.
AC Q27640;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase (EC 1.2.1.3) (Aldehyde dehydrogenase [NAD+]).
GN ALDH.
OS Enchytraeus buchholzi.
OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
OC Tubificina; Enchytraeidae; Enchytraeus.
OX NCBI_TaxID=34589;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96400226; PubMed=8806602;
RA Willum J., Schmitt-Wrede H.-P., Otto A., Wunderlich F.;
RT "Cadmium-detoxification in the earthworm Enchytraeus: specific
expression of a putative aldehyde dehydrogenase.";
RL Biochem. Biophys. Res. Commun. 226:128-134 (1996).
CC -!- FUNCTION: Could be indirectly involved in cadmium-detoxification,
by lowering the intracellular concentrations of aldehydes.
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.

```



```

CC -|- PATHWAY: Ethanol utilization; second step.
CC -|- INDUCTION: By cadmium.
CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X95396; CRA64680.1; -.
CC DR PIR; JC4924; JC4924.
CC DR HSSP; P05091; 1CW3.
CC DR InterPro; IPR002086; Aldehyde_dehydr.
CC DR Pfam; PF00171; aldedh; 1.
CC DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; FALSE_NEG.
CC DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC KW Oxidoreductase; NAD.
CC FT NP_BIND 243 248 NAD (ADP PART) (BY SIMILARITY).
CC FT ACT_SITE 266 266 POTENTIAL.
CC FT ACT_SITE 300 300 POTENTIAL.
CC SQ SEQUENCE 497 AA; 54159 MW; A35A562876808BCC CRC64;

Query Match 25.7%; Score 614; DB 1; Length 497;
Best Local Similarity 32.8%; Pred. No. 3.2e-30;
Matches 158; Conservative 85; Mismatches 208; Indels 30; Gaps 10;

QY 8 YVNGEKSSVN--QTEILSPIDSSLGFPVPMTRREVDHAKMGREALPAAALTVY--- 62
DB 21 FINNEHDSVSGKKFPVNPATEBEKLCVEEGKEDVNKAARQAQFQISFNTMDAS 80
QY 63 ERAQYLHKAADIIRDKBEIATVLAKEISKAY--VTEVVRTADLIRYAAEGIRLST 121
DB 81 DRGRLFLKADLIEHRVQLTLETLDNCKPFAMSYLGDILMAQKVLRYAGFADKIV-- 138
QY 122 ADEGKKWDASTGHKLAVIRQPGVIGVLAIPYNYPNVLSGSKIAPALIGNVVMEKPTQ 181
DB 139 ---GQTIPADGNVFCYTRHEPVGWVGAITPNPFLHLAASKIAPAAGCTLVLPKPAEQ 194
QY 182 GSUSGLVLAKAFABAGLPAAGVNTITGRGSEIGDYIVHEEVNFINFTGTPVGORIGKL 254
DB 195 TPLPALYLASLQKQAGFPAGVINIVFGLGHTAGALTNHPDINKITFTGSTVGQLIQG 254
QY 239 -GKLAGMRPIMLELGKDGAGIVLADADLDNAAKQIVAGAYDSGQRTAKRVLVVEVA 297
DB 255 SGK--TNLKRVTLELGGKSPNIIIPDSOLDYAVEVSHQAIMNMGQVCCAGSRTFVHEDIY 313
QY 298 DELAEKISENVAKLSVGDPPDNATVT--PVIDNSADFTESLVVDARQKAGKELNEFKR-- 354
DB 314 EEFVRSVERAKKRTVGDPPDPKNENGPOVDETQLKLELIESGKTEGAKLEGCGKRLG 373
QY 355 -DGRLLTPGLFDHVTLDMLKAWBEPFPIIIRVKDAEAEVAIAANKSDFGLQSSVTRD 413
DB 374 DKGYFVEPTVFTDVTSSMRVAKKEIFGPVQLIFKQVDEVERIANTSYGLAAAVFTKN 433
QY 414 FQKAFDIANKLEVGTWHNNKTRGPDNF-----PFLGKSGAGVGQIRYISIEAMTNVKS 469
DB 434 IDTALKVNSLEAGIVWNTY-----NHFAQAPGGYKVGSGGREFGHYGLAEFLVKT 488
QY 470 I 470
DB 489 V 489

RESULT 37
DHAI HORSE STANDARD; PRT; 500 AA.
AC PL5437;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1) (ALHDII) (ALDH-E1).
GN ALDH1A1 OR ALDH1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RX MEDLINE=84208025; PubMed=6723662;
RA von Bahr-Lindstrom H., Hempel J., Joernvall H.;
RT "The cytoplasmic isoenzyme of horse liver aldehyde dehydrogenase.
RT Relationship to the corresponding human isoenzyme.";
RL Eur. J. Biochem. 141:37-42(1984).
CC -|- FUNCTION: Binds free retinal and cellular retinol-binding protein-
CC bound retinal. Can convert/oxidize retinaldehyde to retinoic acid
CC (by similarity).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -|- PATHWAY: Ethanol utilization; second step.
CC -|- SUBUNIT: Homotetramer.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
DR PIR; S02302; S02302.
DR HSSP; PS1977; 1EKS.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation.
FT MOD_RES 1 1 ACETYLATION.
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
SQ SEQUENCE 500 AA; 54743 MW; A438E4204BEF7340 CRC64;

Query Match 25.6%; Score 613; DB 1; Length 500;
Best Local Similarity 32.4%; Pred. No. 3.7e-30;
Matches 155; Conservative 89; Mismatches 214; Indels 20; Gaps 8;

QY 8 YVNGEKSSVN--QTEILSPIDSSLGFPVPMTRREVDHAKMGREAL---PAAALTVY 62
DB 23 FINNEHDSVSGKKFPVNPATEBEKLCVEEGKEDVNKAARQAQFQISFNTMDAS 82
QY 63 ERAQYLHKAADIIRDKBEIATVLAKEISKAY--NASVTVVRTADLIRYAAEGIRLST 121
DB 83 ERGRLLYLADLVERDLILATMESNMGGKLFNSAYLMDLGGCLKTLRYCAGWADKIQ-- 140
QY 122 ADEGKKWDASTGHKLAVIRQPGVIGVLAIPYNYPNVLSGSKIAPALIGNVVMEKPTQ 181
DB 141 ---GRTIPSDGNFFTYTRHEPVGWVGQILPNWFLMLWKIAPALSCGNTVVVKPAEQ 196
QY 182 GSUSGLVLAKAFABAGLPAAGVNTITGRGSEIGDYIVHEEVNFINFTGTPVGORIGKL 241
DB 197 TPLSALHVTALKEAGFPGGVNIVPGVGTAGAAISSHMDIKVAFSTGTEVGLIKEA 256
QY 242 AG---MRPIMLELGKDGAGIVLADADLDNAAKQIVAGAYDSGQRTAKRVLVVEVAD 298
DB 257 ACKSNLKRVTLELGGKSPFIFADADLETALEVTHQALFVHQGCCVAAASRLFVEESIYD 316
QY 299 ELAEKISENVAKLSVGDPP--DNAITVPTVIDNSADFTESLVVDARQKAGK---ELNEFKR 354
DB 317 EEFVRSVERAKKRVTLNPLITPGVQSGQFDIKDQYDKILDILIESGKTEGAKLEGCGGPGWN 376
QY 355 DGRLLTPGLFDHVTLDMLKAWBEPFPIIIRVKDAEAEVAIAANKSDFGLQSSVTRDF 414
DB 377 KGYFTQPTVFSNVSDMRVAKKEIFGPVQIQKFKSLDDVIKGNANTTVGLFAGSFTKDL 436
QY 415 QKAFDIANKLEVGTWHNNKTRGPDNFPLGLKSGAGVGQIRYISIEAMTNVKSIVL 472
DB 437 DKAITVSAALQAGTVWV--NCYGVVSAQCPFGGFKMSGNGREMGVEYGFHEYTEVKTIV 493

RESULT 38

```


modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

Query Match 25.4%; Score 609; DB 1; Length 502;
Best Local Similarity 31.2%; Pred. No. 6.6e-30;
Matches 153; Conservative 97; Mismatches 211; Indels 30; Gaps 10;

QY 6 QNYVNGEKSSV--NQIEILSPIDSSGLFVPAMTRVEVDHAKAGREALP-----WAA 58
DB 9 QLFIDGWEPIKKNRIPIINPSNEEIIIGDIPAGSESDIEVAVAAARALKRNGREWAA 68
QY 59 LTVYERAOYLHKAADIERKEIAVLAKESKAYNASVTEVVRTADLIRYA--REGI 116
DB 69 LWSH-RAKYLIRAAIKTEKKOHFVKLETDSGRPRDEAVLDDIDVATCFEYFAGQA 127
QY 114 EGRLSTADEGKMDASTGHKLAVIRROPVIGVILATAPYNPVNLGSKAPALIGNV 176
DB 128 EALDAKQAPVTLPERFKSHV-----RQPIGVVGLISPNWYFLMDTWKIAPALAA 183
QY 174 VNKPTQGSVGLVAKAFABAGLPAQVNTITGRGSEIGYIVVEHEVNFINTGSTP 233
DB 184 TVLKPSELASVTCLEFGEVCEVGLPPGVNLITGLGPDAGAPVSHPDIDKVAFTG 243
QY 234 VGORIGKLAG--MRPIMLEGGKDGAGIVLADLDAKQIVAGAYDYGQRTAKRVL 291
DB 244 TSGKIMASAAQLVKPVLGKSGPVIMFEDIDTAVETWLFVFWTNGQICSAISRLL 303
QY 292 VVEVADELAEKISENVAKLSVGDPP--DNATVTPVIDDNDSDIFSLVVDARQKAGEL- 349
DB 304 VHSIAAEFVDRMVWNTKNIKISDPPEEGCRGLGPVSKGYDKIMKFISTAKSEGATILC 363
QY 350 ----NEFKDGBLLTFLGFOHVTLDKMLAWEPGPILPIIRVKDAEBAVAIAKNSDFGL 405
DB 364 GGRPHLKGKGYFIEPTIITDITTSQIMKEVEFGVICVTKFTEDEALELANDTEYGL 423
QY 406 QSVFTDRFOKAFDANKLEVGTVHNNKTGRGP--DNFPFLGLKSGAGVQGIYSIEA 463
DB 424 AGAVFSKDLERCERVTKALEVGAVWNCQ---PCFVHAPGGVKGSGFRELGEWGIEN 480
QY 464 MTNKSIVLDM 474
DB 481 YLNIKQVTSDI 491

RESULT 40
DHAB BETVU STANDARD; PRT; 500 AA.
AC P28237;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Betaine-aldehyde dehydrogenase, chloroplast precursor (EC 1.2.1.8)
DE (BADH).
OS Beta vulgaris (Sugar beet).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Amaranthaceae; Beta.
CX NCBI_taxid=161934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119215; PubMed=4731961;
RA McCue K.F., Hanson A.D.;
RT "Salt-inducible betaine aldehyde dehydrogenase from sugar beet: cDNA
RT cloning and expression.";
RL Plant Mol. Biol. 18:1-11(1992).
CC -|- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +
CC NADH.
CC -|- PATHWAY: Betaine biosynthesis; last step.
CC -|- SUBUNIT: Homodimer.
CC -|- SUBCELLULAR LOCATION: Chloroplast.
CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR ENBL; X58463; CAA41377.1; -;
DR ENBL; X58462; CAA41376.1; -;
DR PIR; S19135; S19135.
DR HSSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Chloroplast; Transit peptide.
FT TRANSIT 1 7 CHLOROPLAST (POTENTIAL).
FT CHAIN 8 500 BETAIN-ALDEHYDE DEHYDROGENASE.
FT NP BIND 238 243 NAD (ADP PART) (BY SIMILARITY).
FT ACT SITE 260 260 BY SIMILARITY.
FT ACT SITE 294 294 BY SIMILARITY.
FT VARIANT 412 412 E -> D.
SQ SEQUENCE 500 AA; 54720 MW; 512859FBC27B67B8 CRC64;

Query Match 25.4%; Score 608; DB 1; Length 500;
Best Local Similarity 30.3%; Pred. No. 7.5e-30;
Matches 148; Conservative 102; Mismatches 212; Indels 26; Gaps 9;

QY 6 QNYVNGEKSSV--NQIEILSPIDSSGLFVPAMTRVEVDHAKAGREALP-----WAA 59
DB 9 QLFIDGWEPIKKNRIPIINPSNEEIIIGDIPAGSESDIEVAVAAARALKRNGREWAA 68
QY 59 LTVYERAOYLHKAADIERKEIAVLAKESKAYNASVTEVVRTADLIRYA--REGI 116
DB 69 TSGAHRAYLIRAAIKVTERKOHFVKLETIDSGRPFDEAVLDDIDVATCFEYFAGQA 128
QY 117 RLSTADEGKMDASTGHKLAVIRROPVIGVILATAPYNPVNLGSKAPALIGNV 176
DB 129 DAKQAPVTLPERFKSHV-----RQPIGVVGLITPNYPLMATWKIAPALAAAGTAVL 194
QY 177 KPPTQGSVGLVAKAFABAGLPAQVNTITGRGSEIGYIVVEHEVNFINTGSTPVGQ 236
DB 185 KPSELASVTCLEFGEVCEVGLPPGVNLITGLGPDAGAPLAAPVDPVKVAFVTSATGS 244
QY 237 RIGKLAG--MRPIMLEGGKDGAGIVLADLDAKQIVAGAYDYGQRTAKRVLVVE 294
DB 245 KVMASAAQLVKPVLGKSGPVIMFEDIDVDDVQVWTFGCFWNTGQICSAISRLLVHE 304
QY 295 EVADELAEKISENVAKLSVGDPP--DNATVTPVIDDNDSDIFSLVVDARQKAGEL---- 349
DB 305 SIAAEFDRLVKWNTKNIKISDPPEEGCRGLGPVSKGYDKIMKFISTAKSEGATILCGGS 364
QY 350 NEFKDGBLLTFLGFOHVTLDKMLAWEPGPILPIIRVKDAEBAVAIAKNSDFGLQSS 408
DB 365 RPEHLKGKGYFIEPTIITDITTSQIMKEVEFGVICVTKFTEDEALELANDTEYGLASA 424
QY 409 VFTDRFOKAFDANKLEVGTVHNNKTGRGP--DNFPFLGLKSGAGVQGIYSIEAMTN 466
DB 425 VFSKDLERCERVTKALEVGAVWNCQ---PCFVHAPGGVKGSGFRELGEWGIENYLN 481
QY 467 VKSIVLDM 474
DB 482 IKQVTSDI 489

RESULT 41
ROCA STAEF STANDARD; PRT; 514 AA.
AC OSCN04;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 1-pyrioline-5-carboxylate dehydrogenase (EC 1.5.1.12) (p5C
DE dehydrogenase).
GN ROCA OR S82116.
OS Staphylococcus epidermidis.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RK PubMed=12950922;
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
 RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
 RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming
 RT Staphylococcus epidermidis strain (ATCC 12228).";
 RL Mol. Microbiol. 45:1577-1593 (2003).
 CC -|- CATALYTIC ACTIVITY: 1-pyrroline-5-carboxylate + NAD(+) + H(2)O =
 CC L-glutamate + NADH.
 CC -|- PATHWAY: Arginine degradation; third step.
 CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC ROCA subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE016751; AAC05758.1; --
 DR HAMAP; MF 00733; -- 1.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR InterPro; IPR005932; Dipyv5carbox2.
 DR Pfam; PF001171; aldedh; 1.
 DR TIGRFAMs; TIGR01804; BADH; 1.
 DR TIGRFAMs; TIGR01237; Dipyv5carbox2; 1.
 DR TIGRFAMs; TIGR01780; SGADH; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD; Complete proteome.
 FT ACT SITE 286 286 BY SIMILARITY.
 FT ACT SITE 320 320 BY SIMILARITY.
 SQ SEQUENCE 514 AA; 56876 MW; 99EE786D43629FF4 CRC64;

Query Match 25.4%; Score 607.5; DB 1; Length 514;
 Best Local Similarity 31.4%; Pred. No. 8.3e-30;
 Matches 143; Conservative 100; Mismatches 194; Indels 19; Gaps 8;

QY 9 VNGEKSSVQIETILSPDTSSLI-GFVPAMTREEDVHAMKAGREALPAWAALTIVYERAOY 67
 DB 41 INGEKTKYDTNSVNPANTSQLIAKVKATQDDIEKAPESANHAYQSKWWSHKRAEL 100
 QY 68 LHKAADIIBRKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSTSADEGK 127
 DB 101 LLRVAALIRRKEEISAIYVVEAGKPDVAEGDAEGIDFIEYARSMWEL---ADGKPV 157
 QY 128 MDASTGHKAVIRROPVIGVILATAPNYP-VNLSSKIAPALIGNVMPFPPTQGSVSG 186
 DB 158 LDREGEHNYFYK--PIGTGVTTPPNWFFPAMAGTTIAP-VVAGTVLLKPAEDTVLTA 214
 QY 187 LVLAFAFAGAPAGVFNITIGRSIGDYIVHEEVENFNFTGSPVQGIQKLAG--- 243
 DB 215 YKLMEILEBAGLPQGVNVPFGDPKELGDLVDKHDHFTFTFGRAVGTIYERSAVVQ 274
 QY 244 -----MRPIMBLEGKDGAGIVLADADLNAKQIVAGAYDYGQRCQTAIKVLVVEVAD 298
 DB 275 EGQQLKRVIAEWGKDAIVVNDNVDTDLAAEAIVTSAFGFSQKCSARAIVHVDVHD 334
 QY 299 ELAKISENVAKLSVDFPDNATVTVVIDNSADFIESLVVDARQKAKELNEKRD--G 356
 DB 335 EILEKATQLTKLTGNTENTFMGPVINKQFQDKIKNYIEIGKEGKLEGGTDDISG 394
 QY 357 RLLTPGLFHVTLMDKLAWEPFGPIPIIRVKDAEVAIAKNSDFGQSQSVTRDFOK 416
 DB 395 YFIEPTIFSLQASDRIMQEEIFGVPVGIKVKQDFDAIEVANDTYGLTGAVITNREH 454

QY 417 AFDIAKLVGVTVHINK-TGRGPDNFPFLGLKSGG 451
 DB 455 WIKAVNEFDVGNLYLNRGCTAAAVGVYHPPGFGFMSG 490

RESULT 42
 ID XYLCPSEPU STANDARD; PRT; 487 AA.
 AC P43503;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Benzaldehyde dehydrogenase [NAD+] (EC 1.2.1.28).
 GN XYL.
 OS Pseudomonas putida.
 OG Plasmid TOL pWMO, and Plasmid TOL pW53.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RC PLASMID=TOL pWMO;
 RC MEDLINE=95173094; PubMed=7868591;
 RA Inoue J., Shaw J.P., Reik M., Harayana S.;
 RT "Overlapping substrate specificities of benzaldehyde dehydrogenase
 RT (the xylC gene product) and 2-hydroxymuconic semialdehyde
 RT dehydrogenase (the xylG gene product) encoded by TOL plasmid pWMO of
 RT Pseudomonas putida.";
 RL J. Bacteriol. 177:1196-1201 (1995).
 RN [2]
 RN SEQUENCE OF 1-53.
 RC PLASMID=TOL pW53;
 RX MEDLINE=91113163; PubMed=1989592;
 RA Chalmers R.M., Keen J.N., Fawson C.A.;
 RT "Comparison of benzyl alcohol dehydrogenases and benzaldehyde
 RT dehydrogenases from the benzyl alcohol and mandelate pathways in
 RT Acinetobacter calcoaceticus and from the TOL-plasmid-encoded toluene
 RT pathway in Pseudomonas putida. N-terminal amino acid sequences, amino
 RT acid compositions and immunological cross-reactions.";
 RL Biochem. J. 273:99-107 (1991).
 CC -|- CATALYTIC ACTIVITY: Benzaldehyde + NAD(+) + H(2)O = benzoate +
 CC NADH.
 CC -|- SUBUNIT: Homotetramer.
 CC -|- MISCELLANEOUS: Optimal pH is 9.0.
 CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U15151; AAA66218.1; --
 DR EMBL; D63341; BAA09661.1; --
 DR PIR; T47107; T47107.
 DR HGSP; P51977; 1BX5.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF001171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
 KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
 FT NP BIND 232 237 NAD (ADP PART) (BY SIMILARITY).
 FT ACT SITE 254 254 BY SIMILARITY.
 FT ACT SITE 288 288 BY SIMILARITY.
 SQ SEQUENCE 487 AA; 51897 MW; 093CE3E9487AF384 CRC64;

Query Match 25.4%; Score 607; DB 1; Length 487;
 Best Local Similarity 33.3%; Pred. No. 8.4e-30;
 Matches 158; Conservative 91; Mismatches 209; Indels 16; Gaps 9;

QY 2 TVE-----YQNYNGEKSSVQI--EILSPDSSIGFVPAMTREEDVHAMKAGREALPAW 56

```

Db 4 TKEQPIWGVKVSNNWEARGVANVVDPSNGDILGTGVANGEDVDAVNAAKRAQKEW 63
Qy 57 AALTVYERAGLYLHKADIIERDKETATVLAKEISKANASVTEVVTADLIRVAEEGI 116
Db 64 AAIPFSERAAIVKAAEKLKEREYEFADNMVRECGAIRPKGLWEAGIAYEQHQHQA--GL 121
Qy 117 RLSTSDGEGKMDASTGHKLAVIRROPVGVILAIAPYNNPVNLGSGSKIAPALIGNVVMF 176
Db 122 ---ASLPLNGTLFSAVPGMNLQRPVGVGVVIAPNWFLFLAMSVAPALALGNVAIL 178
Qy 177 KPPTQGSVS--GLVLAKAFABAGLPAQVNTITGRSEIGDYIYVEHEVNFNFTGSPVVG 235
Db 179 KPDLQTAVTGGALIAEISDAGMPDGLVHLPG--GADVGSVMVANGIMNISFTGSTQVG 237
Qy 236 QRIKLAG--MRPIMLEGGKDGIVLADADLDNAKQIVAGAVDYSGORCTAIKRLVV 293
Db 238 RLIGKCGRMKKVALEGGNNVHVLVDADLEGAVCAAMGTFLHQGVCMAGRHLVH 297
Qy 294 BEVADELAEKISENVAKLSVGP--FDNATVTPVIDNSADFTLSLVVDARQKAKELNEF 352
Db 298 RDVAQQAELALRAKLVVGDPSNDQVHLGPLINEKQVVRVHALVESAQRAQVLAGG 357
Qy 353 KDGRLLTPLGFLHVTLMKLAWEPEGPILPIIRVKDAEVAIAVKSDPGLQSSVTR 412
Db 358 TYQDRYQATVIMDVKPEMEVFKSEIFGPVAPITVFDSEIEAIELANCSYGLAASHTR 417
Qy 413 DFQKAFDIANKLEGVGVHINKTGRGPDNPFPLGLKSGAGVQ--GIRYSIEAMT 465
Db 418 ALATGLDLAKRLTGMVHINDQPINCEPHVPGMGKSGSGRPGGPASIEFT 471

RESULT 43
DHAM BOVIN
ID DHAM BOVIN STANDARD; PRT; 520 AA.
AC P20000;
DT 01-FEB-1991 (Rel. 17, Created)
DT 28-FEB-2003 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (BC 1.2.1.3) (ALDH
DE Class 2) (ALDH1) (ALDH-E2).
GN ALDH2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90179198; PubMed=1689984;
RA Guan K., Weiner H.;
RT "Sequence of the precursor of bovine liver mitochondrial aldehyde
RT dehydrogenase as determined from its cDNA, its gene, and its
RT functionality.";
RL Arch. Biochem. Biophys. 277:351-360(1990).
RN [2]
RP SEQUENCE OF 92-520 FROM N.A.
RX MEDLINE=89210865; PubMed=2540003;
RA Farres J., Guan K.L., Weiner H.;
RT "Primary structures of rat and bovine liver mitochondrial aldehyde
RT dehydrogenases deduced from cDNA sequences.";
RL Eur. J. Biochem. 180:67-74(1989).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS).
RX MEDLINE=97341232; PubMed=9195888;
RA Steimetz C.G., Xie P., Weiner H., Hurley T.D.;
RT "Structure of mitochondrial aldehyde dehydrogenase: the genetic
RT component of ethanol aversion.";
RL Structure 5:701-711(1997).
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SUBUNIT: Homotetramer.

```

```

CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
DR PIS; S09030; S09030.
DR PDB; 1AG8; 08-OCT-97.
DR PDB; 1A4Z; 08-APR-98.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; 3D-structure.
FT TRANSIT 1 21 MITOCHONDRION.
FT CHAIN 22 520 ALDEHYDE DEHYDROGENASE.
FT NP_BIND 265 270 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 288 288 BY SIMILARITY.
FT ACT_SITE 322 322 BY SIMILARITY.
FT STRAND 42 43
FT STRAND 48 49
FT TURN 52 53
FT STRAND 56 60
FT TURN 62 64
FT STRAND 67 72
FT STRAND 74 74
FT STRAND 76 76
FT TURN 90 90
FT TURN 92 93
FT HELIX 95 98
FT HELIX 101 117
FT TURN 118 118
FT HELIX 119 130
FT HELIX 134 139
FT TURN 140 140
FT HELIX 141 153
FT TURN 154 158
FT STRAND 161 164
FT STRAND 170 178
FT STRAND 181 185
FT HELIX 191 205
FT TURN 206 206
FT STRAND 208 213
FT TURN 215 216
FT HELIX 219 231
FT TURN 232 232
FT TURN 235 236
FT STRAND 238 243
FT HELIX 245 253
FT TURN 254 254
FT TURN 256 257
FT STRAND 260 264
FT HELIX 267 279
FT TURN 280 281
FT STRAND 284 288
FT STRAND 294 297
FT TURN 299 300
FT HELIX 303 315
FT HELIX 316 319
FT TURN 322 323
FT STRAND 327 331
FT HELIX 332 348
FT STRAND 351 351
FT TURN 354 355
FT TURN 357 358
FT STRAND 361 361
FT HELIX 367 383
FT TURN 384 384
FT STRAND 386 389
FT STRAND 393 393
FT STRAND 401 401
FT STRAND 404 407
FT TURN 411 412
FT HELIX 414 417
FT STRAND 424 430
FT HELIX 433 441
FT STRAND 448 452

```

FT STRAND 455 455
 FT HELIX 456 465
 FT STRAND 470 473
 FT TURN 481 482
 FT TURN 490 491
 FT STRAND 492 493
 FT TURN 498 499
 FT HELIX 500 503
 FT TURN 504 505
 FT STRAND 506 514
 SQ SEQUENCE 520 AA; 56708 MW; F920BCFC05B5A5B CRC64;

Query Match 25.4%; Score 607; DB 1; Length 520;
 Best Local Similarity 32.6%; Pred. No. 9e-30;
 Matches 157; Conservative 86; Mismatches 216; Indels 22; Gaps 9;

QY 6 QNYVNGEWSVNO--TEILSPIDDSISLGFVPAMTREVDHAKAGREAL---PAAAL 60
 Db 41 QIFINNEHDAVSKTPTVNPSTGDIHVAGDKADVDRAVKAARAFOLGSPWRMD 100
 QY 61 VYERAOYLKADIIERDKEEATVLAKEISKAYNAS-VTEVVRTADLIRYAAEEGIRLS 119
 Db 101 ASERGGNLLRLADLIERDRTYLALETLDNGKPYIISYLVLDMDVVKCLRYIA----- 153
 QY 120 TSADE-GRWDASTGHKLAVIRROPVGIIVLAIPYNPVNLGSKIAPALIGNVVWFKP 178
 Db 154 GWADKYHGKTIPIIDGDFSYRHEPVGCGQIIPWNPFLMQAKLGPALATGNVVMKV 213
 QY 179 PTOGVSGLVLAAPAFAGLPAAGVNTITGRGSEIGDYIVHEEVEVNFINTGSTPVGQRI 238
 Db 214 AEQTEPLTALYVNLKEAGFPFGVNVNIPGFGPTAGAAIASHEDVDKVAFTSTEVGHIL 273
 QY 239 GKLAG---MRPIMLELGGKAGIVLADADLNAKQIVAGDYDSGORTAKRVLVVEE 295
 Db 274 QVAAGKSNLKRVTLEIGKSPNINSDDMDMVAEQAHFALFNQGCCAGSTFVQSD 333
 QY 296 VADELAEKISENVAKLSVGDPPFNATV--PVIDNSADFTESLVDARQKAGEL---NE 351
 Db 334 IYAEFVERSVARAKSRVGVPPFSRTEQGPQVDETFQKVLGVYIKSGEGLKLCGGA 393
 QY 352 FKRDGRLTPGLFDHVTLMKLAWEPPFGPILPIIRVKDAEAAVAIAKNSDFGLQSSVFT 411
 Db 394 AADRGVFIQPTVFGDLQDGMTIAKEEIFGVVQILKFKSMEEVVGRRANSKYGLAAAVT 453
 QY 412 RDOQAFDANKLEVTHVNNKTRGPNPFGLKSGAGVQGIIRYSIEAMTNKSI 471
 Db 454 KLDKANYLSQALQAGTVWNCYDVFQAS-PFGSYKLSSGRELGEGLQYTEVKTIV 512
 QY 472 L 472
 Db 513 V 513

RESULT 44

ID DHAM LEITA STANDARD; PRT; 498 AA.
 AC Q25417;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH class 2) (P51).
 GN ALDH2.
 OS Leishmania tarentolae (Saurleishmania tarentolae).
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5689;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 10-31.
 RC STRAIN=UC;
 RX MEDLINE=95356798; PubMed=7630384;
 RA Bringaud F., Peris M., Zen K.H., Simpson L.;
 RT "Characterization of two nuclear-encoded protein components of mitochondrial ribonucleoprotein complexes from Leishmania

tarentolae.";
 RL Mol. Biochem. Parasitol. 71:65-79(1995).
 CC -!- FUNCTION: Could have a RNA-binding activity in addition of its catalytic role.
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -!- PATHWAY: Ethanol utilization; second step.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.
 CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z31698; CAAB93503.1; -;
 DR PIR; S43184; S43184.
 DR HSP; P05091; 1CW3.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldehyd_1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; RNA-binding.
 FT TRANSIT 1 9 MITOCHONDRION.
 FT CHAIN 10 498 ALDEHYDE DEHYDROGENASE.
 FT NP_BIND 242 247 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 265 265 BY SIMILARITY.
 FT ACT_SITE 299 299 BY SIMILARITY.
 SQ SEQUENCE 498 AA; 54251 MW; 618D55F6ED5547EC CRC64;

Query Match 25.3%; Score 604.5; DB 1; Length 498;
 Best Local Similarity 31.5%; Pred. No. 1.2e-29;
 Matches 149; Conservative 99; Mismatches 206; Indels 19; Gaps 9;

QY 9 VNGEWSVN--QIEILSPIDDSISLGFVPAMTREVDHAKAGREALPAAALTYERAQ 66
 Db 24 INGRFVAVSGKTFEVNPADEKIVANVAEAKADVDLAVKAAHAFESFMTDCWRN 83
 QY 67 YLHKAADIIERDKEEATVLAKEISKAYNASV-TEVVTADLIRYAAEEGIRLSADE- 124
 Db 84 LMLRLADILEKSKEMALESILDNGKPYEVALNVDVALSVECPYCA--GL-----ADKV 136
 QY 125 GSKMDASTGHKLAVIRROPVGIIVLAIPYNPVNLGSKIAPALIGNVVWFKPPTQSV 184
 Db 137 NGTVPPSGNGLGVKQPIGVCGQIIPWNPFLMAAFKLSPALAMGNTVVLKPAEQPL 196
 QY 185 SGLVAKAFABAGLPAAGVNTITGRGSEIGDYIVHEEVEVNFINTGSTPVGQIRGLA-- 242
 Db 197 TAVRLGEMWMEAGYDPGVNLPFGGATAGSETARHMDVDKIAFTGTAVGHQVQMAAE 256
 QY 243 -GMRPIMLELGGKAGIVLADADLNAKQIVAGDYDSGORTAKRVLVVEEVADELA 301
 Db 257 TNLKKVLELGGKSLTICEDADLEAAEVATRVYFNTGQVCTASRIYVHESYDFV 316
 QY 302 EKISENVAKLSVGDPPFDNA--TVTPVIDNSADFTESLVDARQKAGELNEFKR---DGR 357
 Db 317 SRLKNAEARKVGFNDTGNNGPLVSKQHERVLGYIEDGVKAGATVVTGGKLGDKGY 376
 QY 358 LITPGLFDHVTLMKLAWEPPFGPILPIIRVKDAEAAVAIAKNSDFGLQSSVFTDFOKA 417
 Db 377 FVQPTIFSDVKEDMRICKEEIFGVPVTCVMKYKMDDEVVVKRANDSIYGLAAGICTSMDTA 436
 QY 418 FDIANKLEVTHVNNKTRGPNPFGLKSGAGVQGIIRYSIEAMTNVKS 470
 Db 437 IRYSTYLNAGTVWNTWNNFPC-SNPFGFGKSGIGRELGEVVDVMTPEKAI 488

RESULT 45

FTDH MOUSE
 ID FTDH MOUSE
 AC QBR0Y6;
 STANDARD; PRT; 902 AA.

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 10-formyltetrahydrofolate dehydrogenase (EC 1.5.1.6) (10-FTHFDH).
 GN FTHFD.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Liver;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusica K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + NADP(+) + H(2)O =
 CC tetrahydrofolate + CO(2) + NADPH.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MISCELLANEOUS: CONTAINS TIGHTLY BOUND TETRAHYDROFOLATE
 CC PENTAGLUTAMATE (BY SIMILARITY).
 CC -1- SIMILARITY: In the C-terminal section; belongs to the aldehyde
 CC dehydrogenase family.
 CC -1- SIMILARITY: Contains 1 acyl carrier domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; BC025939; AAH25939.1; -;
 DR EMBL; BC028817; AAH28817.1; -;
 DR EMBL; BC030722; AAH30722.1; -;
 DR EMBL; BC030723; AAH30723.1; -;
 DR EMBL; BC030727; AAH30727.1; -;
 DR EMBL; BC030730; AAH30730.1; -;
 DR MGD; MGI:1340024; Fthfd.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR InterPro; IPR005793; Formyl trans C.
 DR InterPro; IPR002376; formyl trans F.
 DR InterPro; IPR001555; GART AS.
 DR InterPro; IPR006163; Pp bind.
 DR Pfam; PF00171; aldehyd; 1.
 DR Pfam; PF02911; formyl trans C; 1.
 DR Pfam; PF00551; formyl trans F; 1.
 DR PROSITE; PS00373; GART; 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.
 DR PROSITE; PS00075; ACP DOMAIN; 1.
 KW One-carbon metabolism; Oxidoreductase; NADP.
 FT DOMAIN 1 203 GART.
 FT DOMAIN 323 392 ACP CARRIER (ACP).
 FT DOMAIN 417 902 ALDEHYDE DEHYDROGENASE.

FT DOMAIN 106 115 FORMYL-TETRAHYDROFOLATE BINDING
 FT ACT SITE 673 BY SIMILARITY.
 FT ACT SITE 707 BY SIMILARITY.
 SQ SEQUENCE 902 AA; 98709 MW; 9B5526A7FB41909E CRC64;
 Query Match 25.2%; Score 603.5; DB 1; Length 902;
 Best Local Similarity 31.2%; Pred. No. 2.9e-29;
 Matches 155; Conservative 91; Mismatches 214; Indels 37; Gaps 12;
 QY 1 LTKYQYVNVGWE--KSSVNOIEILSPIDSSIGFVPAMTREVVDHAMKAGREALP--AW 56
 Db 418 LQMPYQIFGGEFVDAGAKYTYSTINFGSLVCQVSLAQVSDVDKAAVAAKEAFENGLW 477
 QY 57 AALTYYERAYQLHKAADIIRDKKEIATVLAKIEIKAYNASV--TEVVRTADLIRVAE-- 113
 Db 478 GKINARDGRLLYRLADLMEQHEBELATIEALDAGAVYTLALKTHVMSIQTFRYFAGWC 537
 QY 114 ----EGIRLSTSADEGKMDASTGHKLAVIRQPGVILAIAPYNYPNVLSGSKIAPALIG 170
 Db 538 DKIQGATIPIN-----QARNRNRLTLTKKEPVGCGVIVPNWYPLMLSWKTAACLA 590
 QY 171 GNVVMFKPPTQGSVGLVLAKAFAGLPAQVNTITGRSEIGDYIVBEHEVNFINTG 230
 Db 591 GNTVVKPAQVTPITALKPAELTLKAGIPKGVNVLPGSGSLVQRLSDHPDVRKIGFTG 650
 QY 231 STPVQORIGK---LAGMRPIMLEGGKAGIVLADLADLNAKOIVAGAYDYGORCTAI 287
 Db 651 STEVGKHLKSCALSNNKVKVSLGKSPLIIPADCLNKAVQMGSSVFFNKGNCIAA 710
 QY 288 KRVLVVEVADELAELKISENVAKLSVGDPPDNATVTFVDDNSADFTESLVADAKQKAK 347
 Db 711 GRILFVEDSIHDFQVQKVEEVGKMKIGNPLDRDTNIG--PQNHAEHLKLV-VEYQGVK 767
 QY 348 E-----LNEFKRDRGLLTGLFDHVTLMKLAWEPPFPILPIIRVK--DAEEVAIA 398
 Db 768 EGATLVCGNGVPRPGFPFPQPTVTFVDEHMYIAKEESFGPIIISRFDAGDGDVAVLSRA 827
 QY 399 NKSDPGLQSSVTFDFQKAFDANKLEVGVTHIN--NKTGRGPDNFPLGKGGAGVQG 456
 Db 828 NATEFGLASGVTFTRDKALVYDKLQAGTVFVNTNKTVDAA---PFGFGKSGFGKDL 884
 QY 457 IRYSIEAMTNVKSIVLD 473
 Db 885 GEALNEVRIKTVTFE 901
 RESULT 46
 DHAG HUMAN STANDARD; PRT; 493 AA.
 ID DHAG_HUMAN
 AC P49189;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Aldehyde dehydrogenase, E3 isozyme (EC 1.2.1.3) (Gamma-
 DE aminobutyraldehyde dehydrogenase) (EC 1.2.1.19) (R-aminobutyraldehyde
 DE dehydrogenase).
 DE ALDH9A1 OR ALDH7 OR ALDH9.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT SER-115.
 RC TISSUE=Liver;
 RX MEDLINE=96374830; PubMed=8786138;
 RA Lin S.W., Chen J.C., Hsu L.C., Hsieh C.-L., Yoshida A.;
 RT "Human gamma-aminobutyraldehyde dehydrogenase (ALDH9): cDNA sequence,
 RT genomic organization, polymorphism, chromosomal localization, and
 RT tissue expression.";
 RL Genomics 34:376-380(1996).
 RN [2]
 RP SEQUENCE OF 32-493 FROM N.A., AND PARTIAL SEQUENCE.

RESULT 48
FTDH_RAT
ID_FTDH

```

Db 538 DKTQATPIN-----QARNRNLTUTKKEPVGCVIGVFWNPPLMLLSWKTACLA 590
QY 171 GNVVMEKPPTOGSGVLAKAFAGAGLPAQVNTITGRSGHIGYIVVEHEVNFNTG 230
Db 591 GNTVVKPAQVTPLTALKAELTLKAGIPKGVNVLPGSGSLVGQRLSDHPDVRKIGFTG 650
QY 231 STPVGORIK---LAGMRPIMLBELGQKAGIVLADLDNAKQIVAGAYDYSQCTAI 287
Db 651 STEVGKHMKSALSNVKSLELGGKSLPIIFADCDLANKAVQMGSMVFFNGKNCIAA 710
QY 288 KEVLVVEVADELAELISENVAKLSVGDPEFNATVTPVIDNSADFISLVVDARQKAK 347
Db 711 GRUFVEESHNOFQVQVEEVKKNLGNPLEDNTNG--PQNEAHLKRL-VEYQQRGVK 767
QY 348 E-----LNEFKRDRGLLTPGLFDHVTLDMLKLAWEPPGPIPIIRVK--DAEEAVATA 398
Db 768 EGATLVCGGNQVPRGPFQPTFTDVEDHMYIAKESFGPIMISRFDAGDGDVDAVLSRA 827
QY 399 NKSDPGLQSVTRDFQKADFIANKLEVTYHIN--NKTGRGPDNFPFELGLKSGAGVQ 456
Db 828 NATEFGLASGVTRDINKALVSKLQAGTVFINTYKNTDVA--PPGFKQSGFGKDL 884
QY 457 IRYSIEMTNVKSIVLD 473
Db 885 GEAALNLYLRKIVTFE 901

RESULT 49
DHAB ARATH
ID DHAB ARATH STANDARD; PRT; 501 AA.
AC Q9S95;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Betaine-aldehyde dehydrogenase, chloroplast precursor (EC 1.2.1.8) (BADH)
GN A11G74920 OR F9310.23 OR F25A4.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Millschler J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venner J.C., Davis R.W.
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,

```

```

RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayaishizaki Y., Johnson-Hopson C., Hevan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tanse R., Vayberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
genome.";
RL Science 302:842-846(2003).
CC -|- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +
NADH.
CC -|- PATHWAY: Betaine biosynthesis; last step.
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Chloroplast.
CC -|- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AC013258; AAG51938.1; -.
CC EMBL; AC008263; AAD55284.1; -.
CC EMBL; AY093071; AAM13070.1; -.
CC PIR; H96778; H96778.
CC HSSP; P05091; 1CW3.
CC InterPro; IPR002086; Aldehyde_dehydr.
CC Pfam; PF00171; aldehyd; 1.
CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
CC PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC Oxidoreductase; NAD; Chloroplast; Transit peptide.
FT TRANSIT 1 7 CHLOROPLAST (POTENTIAL).
FT CHAIN 8 501 BETAINE-ALDEHYDE DEHYDROGENASE.
FT NP_BIND 238 243 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 260 260 BY SIMILARITY.
FT ACT_SITE 294 294 BY SIMILARITY.
SQ SEQUENCE 501 AA; 54431 MW; BA8B03C822453B CRC64;
Query Match 24.8%; Score 594; DB 1; Length 501;
Best Local Similarity 31.7%; Pred. No. 5.3e-29;
Matches 154; Conservative 92; Mismatches 214; Indels 26; Gaps 9;
QY 6 QNVYNGEKSSV--NQETLSIDSSISGFVPAMTREVVDHAKAGREALPA-----WAA 58
Db 9 QLFIDGWEREFILKRIPIVNPATEVEIGDIPATTEDDVDVAANRAALSRNKGKDWAK 68
QY 59 LTVYEAQVLHKAADIIRKKEIATVLAKEISKAYNASVTEVVTADLIRYAE--EGI 116
Db 69 APQAVPAKYLRAIAKYNKERTDLAKLEALDCGKPLDEAVMDMDVAGCFEYADLAEL 128
QY 117 RLSTSADEGGKMDASTGHKLAVIRQPIGVILAIAPNYPNLSGSKIAPALIGNVWF 176
Db 129 DAKQKAPVSLPESFKSYVL---KQPLGVVGLITPNYPYELLMAVWKVAPSLAAGCTAIL 184
QY 177 KPPTQSGVGLVAKAFAGAGLPAQVNTITGRSGHIGYIVVEHEVNFNTGSPVQ 236
Db 185 KPSELASTVTELEADI CREVGPPGVNVLVTFGSEAGAPLASHPGVDVKIATGSPATGS 244
QY 237 RIGKLAG--MRPTMLELGGKDGIVLADLDNAKQIVAGAYDYSQCTAIKRVLWVE 294
Db 245 KWTAAQVLKVPVSMELGGKSLPIVFDVDLDKAAEWALFGCFWTNGQICSATSRLLVHE 304
QY 295 EVADELAELKISNVAKLSVGDPPF-DNATVTPVIDNSADFISLVVDARQKAKELN--- 350
Db 305 SIASEFIEKLKVKWSKNIKISDPMEGCRIGFVSVSGQYKILKFTISTAKSEGATILHGS 364
QY 351 --EFKRDGRLITPGLFDHVTLDMLKLAWEPPGPIPIIRVKDAEEAVATAIANKSDFGLOSS 408

```

```
Db 365 RPEHLEKGFIEFTITDVTTSQIWRSEVFGVLCVKTFASEDEAIELANDSHYGAA 424
Qy 409 VFTDFOKAFDANKLEVTGTHNNKTRGP--DNFPLGLKSGAGVQGIYRIEAMTN 466
Db 425 VISNDTERCDRISEAFAGIWINCSQ---PCFTQAPGWGVKRSFGRELGEWGLDNYLS 481
Qy 467 VKSIVL 472
Db 482 VKQVTL 487

RESULT 50
ID _DHAL_AGABI STANDARD; PRT; 500 AA.
AC 074157;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase (BC 1.2.1.3) (ALDDH).
GN ALDA.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Horst H39;
RA Schaap P.J., Muller Y., Visser J.;
RT "Molecular structure and spatial expression of housekeeping genes in
RL mushrooms.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y17825; CAA76875.1; -.
DR HSSP; P51977; 1BXS.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldehyd; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 246 251 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 269 269 BY SIMILARITY.
FT ACT_SITE 303 303 BY SIMILARITY.
SQ SEQUENCE 500 AA; 54395 MW; C4FC58B50855925 CRC64;

Query Match 24.6%; Score 589.5; DB 1; Length 500;
Best Local Similarity 29.5%; Pred. No. 9.9e-29;
Matches 145; Conservative 101; Mismatches 202; Indels 43; Gaps 10;

Qy 8 YNKGKSVNQ--IEILSPIDDSGLGFVPMTRVEVDHAKGREALPAWALTIV--YE 63
Db 25 FINGFVGVKNTTIDVNVNPKGLITIKISEATEADIDIAVEAAHKAFFETWGLNCSGSK 84
Qy 64 RAQYLKADIIERDKETATVLAKESIKAYNASVTEVVRTADLRVAEEGIRLSTAD 123
Db 85 RGDMLTKLAQMEKNIDDLGATELDNGKTF-----LWAKSVLSISITKH 132
Qy 124 EGGKMDASTGH-----KLAVIRQPVGIVLAIPNYPVNLGSKIAPALIGNVNMF 176
Db 133 YAGWADKFGVVIETDEKLTYSRHEPIGVGVGIIIPWNPFLMLLAWKIGPALATGNCIVL 192
Qy 177 KPPTQGSVGLVAKAFAEAGLPAGVFNTITGRGSEIGDIYIIVEEVNFINTFTGTPVGQ 236
```

```
Db 193 KESEETPLSALRCALIQEAGFFPGVNVVTVGGTTGQAISSHWKIDKVAFTGSLVGR 252
Qy 237 RIGKLA---QMRPIMLELGGKDAGIVLADADLNAAKQIVAGAYDSGQRCTAIKRVLIW 293
Db 253 KWEAAAKSNLKNVTLELGGKSPVVIDDADLEQSVNWTAGHLEWNNHGQACCAGTRIFVQ 312
Qy 294 EYVADELAELKISENVAKLVSQDPEQ-NATVTVVIDNSADFIESLVVDAROKGA----- 346
Db 313 EGIYDKFLQKFTDKIKKHKLGGPFGGLGDQGPQVSIQYDRIMSYIESGAEAGATVHVG 372
Qy 347 -KELNEFKRQGRLLITPGLFDHVTLDMLKLAWEPPFGPIIIRVKDAEEAVAJANKSDPGL 405
Db 373 ERHGN-----GYFIQPTIETDTPDMKIVKEEIFPGVAVLKFDGKEVIKQANDSNYGL 428
Qy 406 QSVVTRDFQKAFDIANKLEVGTTHIN--NKTGRGPDNPPFLGLKSGCAGVQGIYSIEA 463
Db 429 AAASFQDINKAETATAHAFKAGTAWUNCANTIDAG---VPPGGYKQSGIGIGELGEYALHN 485
Qy 464 MTNVKSIIVLDM 474
Db 486 YTNVKAHEVNL 496
```

Search completed: April 13, 2004, 14:21:27
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:17:48 ; Search time 21 Seconds
(without alignments)
2175.761 Million cell updates/sec

Title: US-09-868-195-12
Perfect score: 2393
Sequence: 1 LIKEQYVNGEKSSVNOI.....GIRYSEAWTVKSIIVLDMK 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR 78:*

1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1812	75.7	474	2 E95129	hypothetical prote
2	1802	75.3	474	2 D98000	glyceralddehyde-3-p
3	1702	71.1	475	2 A57151	glyceralddehyde-3-p
4	1383	57.8	482	2 H97348	NADP-dependent gly
5	1244	52.0	481	2 B83929	NADP-dependent gly
6	1168.5	48.8	496	2 F84634	hypothetical prote
7	1164.5	48.7	498	1 S43833	glyceralddehyde-3-p
8	1143.5	47.8	496	1 S43832	glyceralddehyde-3-p
9	868.5	36.3	475	2 F82900	NADP-dependent gly
10	763	31.9	488	2 G69752	aldehyde dehydroge
11	753.5	31.5	463	2 B64476	NADP-dependent gly
12	752	31.4	470	2 F90347	hypothetical prote
13	750	31.3	470	2 D90323	hypothetical prote
14	741.5	31.0	455	2 H69230	NADP-dependent gly
15	741.5	31.0	489	2 A83033	probable aldehyde
16	737	30.8	476	2 A70318	aldehyde dehydroge
17	736.5	30.8	491	2 T46684	p-hydroxybenzaldehy
18	735	30.7	482	2 B93867	NADP-dependent gly
19	726	30.3	482	2 F65045	succinate-semialde
20	721.5	30.2	488	1 I39769	aldehyde dehydroge
21	720	30.1	488	2 A11546	succinate semialde
22	718	30.0	488	2 A11168	succinate semialde
23	717	30.0	482	2 B91069	succinate-semialde
24	717	30.0	482	2 D85913	succinate-semialde
25	713.5	29.8	480	2 C84250	glyceralddehyde-3-p
26	713	29.8	501	2 T44939	glyceralddehyde-3-p
27	710.5	29.7	489	2 C84401	aldehyde dehydroge
28	710.5	29.7	498	2 AD2277	aldehyde dehydroge
29	697.5	29.1	468	2 C83774	succinate-semialde

30 682.5 28.5 485 2 B83901 aldehyde dehydroge
31 682 28.5 482 2 AE0839 succinate-semialde
32 682 28.5 496 2 T44987 aldehyde dehydroge
33 673 28.1 485 2 C69584 aldehyde dehydroge
34 669.5 28.0 470 2 A87547 vanillin dehydroge
35 668 27.9 483 2 D83613 succinate-semialde
36 666.5 27.9 524 2 E75574 aldehyde dehydroge
37 666 27.8 509 1 S14629 aldehyde dehydroge
38 664.5 27.8 495 2 C89778 hypothetical prote
39 664 27.7 477 2 AF3469 aldehyde dehydroge
40 661.5 27.6 509 2 E90504 hypothetical prote
41 661.5 27.6 509 2 E96825 hypothetical prote
42 661 27.6 517 1 A40872 aldehyde dehydroge
43 660 27.6 483 2 B49343 salicylaldehyde de
44 658.5 27.5 490 1 S15181 betaine-aldehyde d
45 658 27.5 501 1 DEHUE1 aldehyde dehydroge
46 656 27.4 485 2 C95964 probable aldehyde
47 655 27.4 487 2 AD3300 succinate-semialde
48 653.5 27.3 490 2 B85524 NADP-dependent bet
49 653 27.3 477 2 A95990 probable dehydroge
50 651.5 27.2 485 2 H72562 probable aldehyde

ALIGNMENTS

RESULT 1

E95129
hypothetical protein Sp119 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: E95129
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; WUID:21357209; PMID:11463916
A:Accession: E95129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75230.1; PID:gl4972596; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 75.7%; Score 1812; DB 2; Length 474;
Best Local Similarity 74.1%; Pred. No. 7.4e-110;
Matches 349; Conservative 57; Mismatches 65; Indels 0; Gaps 0;

QY 5 YQYVNGEKSSVNOIETILSDSSIGFVPANTREYDHWAKAGREALPAWALTVER 64
DB 4 YQYVNGEKSSVNOIETILSDSSIGFVPANTREYDHWAKAGREALPAWALTVER 63
QY 65 AYLHKAADIIRDKKEITATVLAKEISKAYNASVTEVVTADLIRYAAEGIRLSTADE 124
DB 64 AYLHKAADIIRDKKEITATVLAKEISKAYNASVTEVVTADLIRYAAEGIRLSTADE 123
QY 125 GGMKDASTGHKLAVIRPQVGLVLAIPVNYPNVGGKIAPALIGNVNMPKPTQGSV 184
DB 124 GGFPEATSKNLAVRPFVGLVLAIPVNYPNVGGKIAPALIGNVNMPKPTQGSV 183
QY 185 SGLVLAFAEAGLPAGVNTITGRSGEIGDYIIEHKEVFNFTGTPVGORIGKLAGM 244
DB 184 SGLVLAFAEAGLPAGVNTITGRSGEIGDYIIEHKEVFNFTGTPVGORIGKLAGM 243
QY 245 RPIMLEGGKDAQVIADADLDNAKQIVAGADYSGORCTAIKRVIVVEEVADEAEKI 304
DB 244 RPIMLEGGKDAQVIADADLDNAKQIVAGADYSGORCTAIKRVIVVEEVADEAEKI 303

A:Residues: 1-482 <KUR>
 A:Cross-references: GB:AB001437; PIDN:AAK81579.1; PID:gl5026759; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3657
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 57.8%; Score 1383; DB 2; Length 482;
 Best Local Similarity 57.0%; Pred. No. 4.7e-82;
 Matches 269; Conservative 86; Mismatches 115; Indels 2; Gaps 2;

QY 5 YQNYNGW-KSSVNO-IETLSPIDSSLGFPVPMATREEDVHAMKAGREALPAAALTY 62
 DB 11 YKMLFDGKWSKNTKIETHSPYDGLIGKQVALSKEEVEIDFKSSRTAKKWTGPIN 70
 QY 63 ERAQYLHKAADIIRDEKEEATVLAKEISKAYNASVTVVRADLIRVAEEGIRLSTSA 122
 DB 71 ERARIMKAAADILDNNAEYIAKILSNEIAKDLKSLSEVKRTADFRPTANEETHMEGEA 130
 QY 123 DEGGKMDASTGHKLAIVRROPVIGVLAIPYVNVNLSGKIAPALIGNVVMEKPEPTOG 182
 DB 131 INSDNFPQSKDKLSIVERFVLGIVLAISPENPVNLSGKVPALIGNSVVKESTTG 190
 QY 183 SVSGLVLAFAEAGLPAQVNTITGRGSEIGDYIVHEBEVNFINTGSTPVGQRIKLA 242
 DB 191 AISALHLEIFNAAGLPAGVINTVTKGSEIGDYLITHEEVNFINTGSSAVGKHISKIA 250
 QY 243 GMRPIMLEGGKAGIVLADADLDNAKQIVAGAYVSGORCTAIKRVLVVEEVADELA 302
 DB 251 GMPVMVLEGGKAAVLEADANLETTAKSIVSGAYGYSQRCCTAVKRVLMVMDKVADELVE 310
 QY 303 KISENVAKLSVGDPEPDNATVTPVIDNSADFIESLVVDAROKGAKELNEFKRGLLITPG 362
 DB 311 LVTKVKEIKVGNPFDDVTITPLIDNKAADYVQTLIDDALEKAGATLVGNKRENLMYPT 370
 QY 363 LFHDVTLDMKLAWEPEPGPILPIIRVKDAEBAVAIAANKSDFGLQSVFTRDFQKAFDIAN 422
 DB 371 LFONVTDNRKIAWEPEPGPILPIIRVKSDMAEIAELANSEYGLQSAVTENHDAFYIAN 430
 QY 423 KLEVGVTHNNKTRGPDNPFPLGLKSGAGVQGIKIRYSIEAMTVKSVILDM 474
 DB 431 KLDVGVTVQNNKPERGPDHFPFLGLTKSGMGVQGIKIRYSIEAMTRKSVILNL 482

RESULT 5
 E83929
 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase gapN [imported] - Bacillus halodurans
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

A:Accession: E83929
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E83929
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-481 <STO>
 A:Cross-references: GB:AP001514; GB:BA000004; NID:gi10174613; PIDN:BA005956.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: gapN
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 52.0%; Score 1244; DB 2; Length 481;
 Best Local Similarity 51.4%; Pred. No. 4.7e-73;
 Matches 240; Conservative 95; Mismatches 130; Indels 2; Gaps 1;

QY 10 NGEWKS--VNQTEILSPIDSSLGFPVPMATREEDVHAMKAGREALPAAALTYVYRAQY 67
 DB 15 NGEWESRTERISISAPASGVALGSIPALSQEVDNIAIQAKDAQKIWKTRPIHVDL 74
 QY 68 LHKAADIIRDEKEEATVLAKEISKAYNASVTVVRADLIRVAEEGIRLSTSADEGGK 127

DB 75 LYAWADLLSERKEIITGELIMHEVAKPKSAIGVSRRTADIIRHTADEALRLNGETLKGQD 134
 QY 128 MDASTGHKLAIVRROPVIGVLAIPYVNVNLSGKIAPALIGNVVMEKPEPTOGSVSL 187
 DB 135 PKGSSKKIALVEREPVIGVLAISPENPVNLSGKIAPALYVTVVFKATGSSLSGI 194
 QY 188 VLAFAEAGLPAQVNTITGRGSEIGDYIVHEBEVNFINTGSTPVGQRIKLAGMRPI 247
 DB 195 KWVEALADAGAPGEGIQVTVGRSGVIGDHLVHEHFDIMITFTGGTITGERISERAKMIPV 254
 QY 248 MLEGGKMDAGIVLADADLDNAKQIVAGAYVSGORCTAIKRVLVVEEVADELAELKISEN 307
 DB 255 VLEGGKMDPAIYLDADADLKLTASQIVSGAFYSQRCCTAIKRVFQDSVADQLVANKEL 314
 QY 308 VAKLSVGDPEPDNATVTPVIDNSADFIESLVVDAROKGAKELNEFKRGLLITPGFDHV 367
 DB 315 VEQUTVSGPEDDADITPVIDEKSAFIQGLIDDALENGATLLSGNKROGNULLFTLDDV 374
 QY 368 TDMKLAWEPEPGPILPIIRVKDAEBAVAIAANKSDFGLQSVFTRDFQKAFDIANKLEV 427
 DB 375 TPANVAVEEPGPVLPPIIRVKDANEALSLSNOSDYGLQASIFTKDTDRAINIKHLEV 434
 QY 428 TVHNNKTRGPDNPFPLGLKSGAGVQGIKIRYSIEAMTVKSVILDM 474
 DB 435 TVHINAKTERGPDHFPFLGLVKSGVQGIKPSLLSMTRERVTLML 481

RESULT 6

F84634
 hypothetical protein At2g24270 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: F84634

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: F84634
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-496 <STO>
 A:Cross-references: GB:AE002093; NID:g4115387; PIDN:AAD03388.1; GSPDB:GN00139
 A:Map position: 2
 A:Gene: At2g24270
 A:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 48.8%; Score 1169.5; DB 2; Length 496;
 Best Local Similarity 50.3%; Pred. No. 3.8e-68;
 Matches 240; Conservative 84; Mismatches 142; Indels 11; Gaps 5;

QY 5 YQNYNGWK--SSVNQTEILSPIDSSLGFPVPMATREEDVHAMKAGREALPAAALTYV 62
 DB 16 YKYADGEWKTSSSGKSAVIMNPATRTKTYQVQACTQEEVNVNVELAKSAQKSWAKTILW 75
 QY 63 ERAQYLHKAADIIRDEKEEATVLAKEISKAYNASVTVVRADLIRVAEEGIRLSTSA 122
 DB 76 KRAELHKAALIKDNKAPMAESLVKEIAKPAKDSVTVVRSGDLISYCAEGBVRI--L 132
 QY 123 DEGGKMDASTGH--KLAIVRROPVIGVLAIPYVNVNLSGKIAPALIGNVVMEK 177
 DB 133 GEGKFLSDSPGNDRTKYCLTSKIPLGVLAIPTFPNPNVNLAVSLAPALAGNSVLVK 192
 QY 178 PPTQSSVSLVLAKAFAEAGLPAGVNTITGRGSEIGDYIVHEBEVNFINTGSTPVGQR 237
 DB 193 PPTQGAVALHMHVECHLAGFPKGLISICITKGSEIGDFLTNHPAVNCISFTGG-DTGIS 251
 QY 238 ICKLAGMRPIMLEGGKMDAGIVLADADLDNAKQIVAGAYVSGORCTAIKRVLVVEEVA 297
 DB 252 ISKAGMPLQMGELGKDAACIVLDDADLDLVAASNIIKGFYSQRCCTAVKVLVWESVA 311

QY 298 DELAEKISENVAKLSVGDFFDNATVTPVIDNSADFISSLVVDAROKGAKELNEFKRDR 357
 DB 312 DELVEKAKVAKLTGVGPEENSDITAVSSANFIEGLVMDAKEKATFCQEKREGN 371
 QY 358 LITPGLFHVTHLMDKLAWPEEPFGLPIIRVKDAEAAVAIAANKSDFGLOSSVFTDFQKA 417
 DB 372 LIWPLLDNVRPDMRIAWEEPFGVPVLRINSVEEGIHNCNAGNFGLOGCVFTKDINKA 431
 QY 418 FDIANKLEVTGVTHINNTKGRGPNFPFLGKSGAGVQGIIRYSIEAMTNVKSIVLDM 474
 DB 432 ILISDAMETGTVQINSAPARGDPHFPGQLKDSGSGQGVTSINLMTKVTVINL 488
 RESULT 7
 S43833
 glyceraldehyde-3-phosphate dehydrogenase (NADP) (EC 1.2.1.9) - maize
 C:Species: Zea mays (maize)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
 C:Accession: S43833
 R:Habenicht, A.; Hellman, U.; Cerff, R.
 J. Mol. Biol. 237, 165-171, 1994
 A:Title: Non-phosphorylating GAPDH of higher plants is a member of the aldehyde dehydrogenase family
 A:Reference number: S43832; MUID:94180387; PMID:7545914
 A:Accession: S43833
 A:Molecule type: mRNA
 A:Residues: 1-498 <HAB>
 A:Cross-references: EMBL:X75327
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: NADP; oxidoreductase
 F:56-320/Domain: aldehyde dehydrogenase homology <ALDD>
 F:266,300/Active site: Glu, Cys #status predicted
 Query Match 48.7%; Score 1164.5; DB 1; Length 498;
 Best Local Similarity 49.0%; Pred. No. 76-68;
 Matches 234; Conservative 88; Mismatches 143; Indels 13; Gaps 5;
 QY 5 YQNVNNGKWSVNV--QIEILSPIDDSGLGFPVAMTRREEDVHAMKAGREALPAWALTY 62
 DB 18 YRYADGHWRTSASGKVALNFTTRKTYQYRQACTQEEVNKAMDAAKVAQKAWARTPLW 77
 QY 63 ERAQYLHKAADIIRDEKEETATVLAKEISKAYNASVTEVVRADLIRYAEIGIRLSTSA 122
 DB 78 KRAVLHKAADILKEHKAIAECLVKEIAKPAKDANSEVVRSGDIVSYTAEGRVILGE- 136
 QY 123 DEGKM---DASTG---KLAIVRQPVGIVLAIAPVNPVNLSSKIAPALIGNVNMF 176
 DB 137 ---GKLVSDFSPPGNERNKYCLSKPLGLVLAIPFPNPANLASKIGIPALLAGNALVL 193
 QY 177 KPPTQSVSLVLAKAFAGVAGVNTITGRSGEIGDYIVVEHEVNFINTGSTPVQ 236
 DB 194 KPPTQGAVALHWHVCHFLAGFPKGLISCVTSGKSGEIGDFTLTHPGVNCISFTGG-DTGI 252
 QY 237 RIGKLAMPIMLELGGKDAIGVLAADLNNAAQIVAGAYDSQRCATKRVLVVEVA 296
 DB 253 AISKKAGWVPLQMEILGGKDAICVLEADLNLVSANIVKGSFYSQRCATKRVLVVMEI 312
 QY 297 ADELAEKISENVAKLSVGDFFDNATVTPVIDNSADFISSLVVDAROKGAKELNEFKR 356
 DB 313 ADAVQKNVAKLAKLVGPPEDSDITPVVSSANFIEGLVMDAKEKATFCQEKREGN 372
 QY 357 RLITPGLFHVTHLMDKLAWPEEPFGLPIIRVKDAEAAVAIAANKSDFGLOSSVFTDFQKA 416
 DB 373 LIWPLLDNVRPDMRIAWEEPFGVPVLRINSVEEGIHNCNAGNFGLOGCVFTKDINKA 432
 QY 417 AFDIANKLEVTGVTHINNTKGRGPNFPFLGKSGAGVQGIIRYSIEAMTNVKSIVLDM 474
 DB 433 ALLISDAMETGTVQINSAPARGDPHFPGQLKDSGSGQGVTSINLMTKVTVINL 490
 RESULT 8
 S43832
 glyceraldehyde-3-phosphate dehydrogenase (NADP) (EC 1.2.1.9) - garden pea

C:Species: Pisum sativum (garden pea)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
 C:Accession: S43832
 R:Habenicht, A.; Hellman, U.; Cerff, R.
 J. Mol. Biol. 237, 165-171, 1994
 A:Title: Non-phosphorylating GAPDH of higher plants is a member of the aldehyde dehydrogenase family
 A:Reference number: S43832; MUID:94180387; PMID:7545914
 A:Accession: S43832
 A:Molecule type: mRNA
 A:Residues: 1-496 <HAB>
 A:Cross-references: EMBL:X75327
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: NADP; oxidoreductase
 F:54-318/Domain: aldehyde dehydrogenase homology <ALDD>
 F:264,298/Active site: Glu, Cys #status predicted
 Query Match 47.8%; Score 1143.5; DB 1; Length 496;
 Best Local Similarity 48.8%; Pred. No. 1.6e-66;
 Matches 233; Conservative 85; Mismatches 148; Indels 11; Gaps 5;
 QY 5 YQNVNNGKWSVNV--QIEILSPIDDSGLGFPVAMTRREEDVHAMKAGREALPAWALTY 62
 DB 16 KYIADGHWRTSASGKVALNFTTRKTYQYRQACTQEEVNKAMDAAKVAQKAWARTPLW 75
 QY 63 ERAQYLHKAADIIRDEKEETATVLAKEISKAYNASVTEVVRADLIRYAEIGIRLSTSA 122
 DB 76 KRAELLHKAADILKEHKAIAECLVKEIAKPAKDANSEVVRSGDIVSYTAEGRVRI---L 132
 QY 123 DEGKM---DASTG---KLAIVRQPVGIVLAIAPVNPVNLSSKIAPALIGNVNMF 177
 DB 133 GEGKFLVSDSFPGNERNKYCLSKPLGLVLAIPFPNPANLASKIGIPALLAGNALVL 192
 QY 178 PPQTQSVSLVLAKAFAGVAGVNTITGRSGEIGDYIVVEHEVNFINTGSTPVQ 237
 DB 193 PPQTQGAVALHWHVCHFLAGFPKGLISCVTSGKSGEIGDFTLTHPGVNCISFTGG-DTGI 251
 QY 238 IGLKAMPIMLELGGKDAIGVLAADLNNAAQIVAGAYDSQRCATKRVLVVEVA 297
 DB 252 ISKKSGMIPLQMEILGGKDAICVLEADLNLVAANIIGKGSFYSQRCATKRVLVVMEI 311
 QY 298 DELAEKISENVAKLSVGDFFDNATVTPVIDNSADFISSLVVDAROKGAKELNEFKRDR 357
 DB 312 DALVEKVKVAKLVGPPEDSDITPVVSSANFIEGLVMDAKEKATFCQEKREGN 371
 QY 358 LLTPGLFHVTHLMDKLAWPEEPFGLPIIRVKDAEAAVAIAANKSDFGLOSSVFTDFQKA 417
 DB 372 LIWPLLDNVRPDMRIAWEEPFGVPVLRINSVEEGIHNCNAGNFGLOGCVFTKDINKA 431
 QY 418 FDIANKLEVTGVTHINNTKGRGPNFPFLGKSGAGVQGIIRYSIEAMTNVKSIVLDM 474
 DB 432 IMISDAMESGTVQINSAPARGDPHFPGQLKDSGSGQGVTSINLMTKVTVINL 488
 RESULT 9
 F82900
 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase U0362 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: F82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mini
 A:Reference number: A82870
 A:Accession: F82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-475 <GLA>
 A:Cross-references: GB:AE002133; GB:AF222894; NID:96899339; PIDN:AAF30771.1; GSPDB:GN0011
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: gapN; U0362
 A:Genetic code: SGC3

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 Query Match 36.3%; Score 868.5; DB 2; Length 475;
 Best Local Similarity 39.7%; Pred. No. 9.7e-49;
 Matches 190; Conservative 95; Mismatches 177; Indels 17; Gaps 5;

QY 5 YQYVNGWKSQVNVQIIEISPDSDSLGFPVPMTR-EVDHAKAGREALPAAALTYVE 63
 DB 3 YKTLINGAFVDAKEKLPVNPNSNQIAYVPNIHNEINTIFENAHIAVYKFPDPIKY 62

QY 64 RAQYLHKAADIIEERDKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLS 122
 DB 63 RCDLLKADLDEHKEIAQIISTEIAKGLKSLIEVSADYLRETFEYQKLMQKEPI 122

QY 123 --DEGKMDASTGH----KLAVIRQPVGIIVAIAPVNPVNLSSKIAPALGNNVWF 176
 DB 123 IFDE-----TVHHVKNVATYRIPVGVVLAICPNYPINLLASKLAPALVSGNSLYV 175

QY 177 KPTQSGVSLVLAKAFABAGLPAGVNTITGRSEIGDYVIEHEEVNFNFTGSPVGO 236
 DB 176 KPSTQSLIGIRISELVHEVGFPGVNVNCLTEARITGDLVNTKVIKALISFTGGPKVN 235

QY 237 RIGKLAGMPIMLELGGKAGIVLADADLNAAKQIVAGAYDSQRCATIKRYLVVEEV 296
 DB 236 HIATETSKISLVLELGGKDPALVDADDFELAAEIVKYGFGSGRCATIKRVFVSHN 295

QY 297 ADELAKISENVAKLSVGDPPDNATVTPVIDNSADFIESLVVDARQGA--KELNEFKR 354
 DB 296 HDLLVNLINKVDALTVLGPQNPIITPLINSNLSKYNLSLVEDAIKKGALVHOKIYNE 355

QY 355 DGRLLPGLFDHVTLDMLKAWEPFGPIPIIRVKDAEEAVATANKSDFGLOSSVFTDF 414
 DB 356 KNNLLPLVDNVTMKVAVWEFPGPIPIITNSIQEALDINOSQYGLQACIFITNY 415

QY 415 OKAPDIANKLEVTGHVHNNKTGRGPNFPFLGLKSGAGVQGRYISIAMTNKSVILD 473
 DB 416 ASIQLALQESGTININKSSRGPDILPFGVKDQSGFGVQGVVDAILSMITKIGIIN 474

RESULT 10
 G69752
 aldehyde dehydrogenase homolog ycbD - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: G69752
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 V. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: G69752
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-488 <KUN>
 A:Cross-references: GB:Z99105; GB:ALJ009126; NID:G2632457; PIDN:CAB12041.1; PID:G2632533
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ycbD
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 F:49-309/Domain: aldehyde dehydrogenase homology <ALD>

Best Local Similarity 36.9%; Pred. No. 6.9e-42;
 Matches 170; Conservative 94; Mismatches 179; Indels 18; Gaps 6;

QY 5 YQYVNGW--KSSVNVQIIEISPD--DSSLGFPVPMTR-EVDHAKAGREALPAAALTYV 61
 DB 10 YLNFINGEWFVKSQSDVMKVPADYNDIVGVQNSAEDVERAVTAANEAKTARLKTG 69

QY 62 YERAOYLHKAADIIEERDKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLS 121
 DB 70 AERQYLYKTADIMQRLBEELACATRENGKTLPEAKGETARGTALIRYVYAGEGRKT-- 127

QY 122 ADEGKMDASTGHKLAVIRQPVGIIVAIAPVNPVNLSSKIAPALGNNVWFKEPTQ 191
 DB 128 ---GDVIPSTDKDALMFTTRVPLGVGVISPNWFFVPIPKMAPALVYVNTVWIKPATE 184

QY 182 GSVGLVLAKAFABAGLPAGVNTITGRSEIGDYVIEHEEVNFNFTGSPVGOIRGKL 241
 DB 185 TAVTCAKTIACFEAGLPAGVNTITGRSEIGDYVIEHEEVNFNFTGSPVGOIRGKL 244

QY 242 AGMR--PIMLELGGKAGIVLADADLNAAKQIVAGAYDSQRCATIKRYLVVEEVADE 299
 DB 245 ALARGAKVQLEMGKNPVI VADADLEAAEAVITGAPRSTGKCTATSKRVIVQSGIYER 304

QY 300 LAEKISENVAKLSVGDPP--DNATVTPVIDNSADFIESLVVDARQKAGEL-----NE 351
 DB 305 FKEKLLQRTKDIITIGSLKEDVWVGPTASKNQDLNCLSYIEKKGQEGASLLIGGEKLENG 364

QY 352 FKRGRLTGLFDHVTLDMLKAWEPFGPIPIIRVKDAEEAVATANKSDFGLOSSVFT 411
 DB 365 KYQNGYVQPAIFDNVTSEMTIAQEEIFGVPVALLKVDSTIEALINADVKEGLSASFT 424

QY 412 RDPQKAFDIANKLEVTGHVHNNKTGRGPNFPFLGLKSGGA 452
 DB 425 ENIGRMLSFIDEIDAGLVINAESAGVELQAPFGGKQSS 465

RESULT 11
 B64476
 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.1-) - Methanococcus jann
 C:Species: Methanococcus jannaschii
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Aug-1998
 C:Accession: B64476
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
 rson, J.D.; Sadow, P.W.; Hauna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurs, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: B64476
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-463 <BUL>
 A:Cross-references: GB:U67581; GB:L77117; NID:G1592052; PID:G1592060; TIGR:MJ1411; PID:G1
 C:Genetics:
 A:Map position: FOR1371808-1373199
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: NADP; oxidoreductase

Query Match 31.5%; Score 753.5; DB 2; Length 463;
 Best Local Similarity 36.9%; Pred. No. 2.6e-41;
 Matches 174; Conservative 95; Mismatches 184; Indels 19; Gaps 8;

QY 8 YQYVNGWKSQVNVQ--IETLSPDSDSLGFPVPMTR-EVDHAKAGREALPAAALTYV 65
 DB 2 FIDGKW---INEDMOVINPYSYLVKIPALSRKEAIDAETAEKYVMKNLPITTKRY 58

QY 66 QYLHKAADIIEERDKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSISADBG 125
 DB 59 NILMIAQIKKEKEELAKILADAGKIKQARVVERSISGTFKLAA---FYVKEHREDEV 115

QY 126 GKMDASTGHKLAVIRQPVGIIVAIAPVNPVNLSSKIAPALGNNVWFKEPTQSGVS 185

Db 116 IPSD---DELIFTEPVPVIGVITPFPNPLNSAKIAPALATGNTVIVHHPSSKAPLV 171
Qy 186 GLVLAK-----AFAEAGLPAGVFNITIGRSEIGDYIVVEHEVNFNFTGSPVQSGIKL 241
Db 172 CIELAKIENALKNVPLVYLLTGAGEVVGDEIVVNEKNM-SFTGSSKVGELITKK 231
Qy 242 AGKFPIMLELGGKAGIVLADADLNAKQIVAGAYDSQRCATKRVLVVEEVADELA 301
Db 232 AGFKKALELGGVNPVILKADLNKAVNALIKSFYIAGQVCISVGMILVDESIAKFI 291
Qy 302 EKISENVAKLSVGDPDFNAT--VTPVIDDMSADFIESLVDARQKGAKEINFEKDEGLIT 360
Db 292 EMFVYKAKVLNNGNPLDEKTDVGLISVEHAEWKEVKAIDEGGKLLGKEDKALFY 351
Qy 361 PGLFDHVTLMKLAWEPPGPIILPIIRVKDAEAAVANKSDFGLOSSVFTDFQKAFDI 420
Db 352 PTILE-VDRONILCKTETFAPIPIRTNE-EEMDIANSTGYLHSAIFNDINKSLKF 409
Qy 421 ANKLEVTGHNNKTRGPDNFPPLGLKGGAGVQIRYSIEAMTNVKSIVL 472
Db 410 AENLEFGVGVINDSLSFRQDNMPGGVYKSGLGRGVYKAEWMSNIXITII 461

RESULT 12
P90347
hypothetical protein gapN-2 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
A:Accession: F90347
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: F90347
A:Molecule type: DNA
A:Residues: 1-470 <KUR>
A:Cross-references: GB:AE006641; NID:gl3815099; PIDN:AAK42037.1; GSPDB:GN00155
C:Genetics:
A:Gene: gapN-2
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 31.4%; Score 752; DB 2; Length 470;
Best Local Similarity 36.5%; Pred. No. 3.4e-41;
Matches 172; Conservative 96; Mismatches 187; Indels 16; Gaps 7;

Qy 16 SVNQTIELSPIDSSSLGFPVAMTREVVDHAKAGREALPAAALTVYERAYQLHKAADI 75
Db 3 SLEIVSRPSNLKAVITVKRMKDEVGEIEEAYKGPILSRMPLYKRTAILRKISBIL 62

Qy 76 ERDKEIATVLAKESKAYNASVTVETADLIRYAAEE-GIRLSTADSGKKMA--- 130
Db 63 EREQEARTLAMEAGKPIRDSRVETLRASLFRHAAEEVGWLE---GKNYVDAYEYYP 119

Qy 131 -STGKHLAVIRQPGVILAIAPYVNLGSKIAPALIGNVNMFKPTQGSVGLV 189
Db 120 PGNENRVLSTREPIGVVTTALFPNFPINSFAHKVAPALAVGNSVVKPSINTPLAAVEM 179

Qy 190 AKAFAGLPAGVFNITIGRSEIGDYIVVEHEVNFNFTGSPVQSGIKLA---GMRP 246
Db 180 KKLIVAGLPDSAVRVVTVGYSSEIGDEIITHTPLVGLITLTGTQTGLKIAKASISLGR- 238

Qy 247 IMLELGGKAGIVLADADLNAKQIVAGAYDSQRCATKRVLVVEEVADELAEKISE 306
Db 239 IIMELGSDPIIILEDANIERASSIAVRARFEYAGQNCNAGKRIIVROEVYDFKFAFNE 298

Qy 307 NVAKLSVGDPDFNAT--VTPVIDDMSADFIESLVDARQKGAKE--LNEFKRDGRLLTPGL 363
Db 299 KAKALKVGFPLDETITDVGFINKESENLSNVLDEAKVKGRVEILNKPESGSPFLTM 358

Qy 364 FDHVTLMKLAWEPPGPIILPIIRVKDAEAAVANKSDFGLOSSVFTDFQKAFDIANK 423

Db 359 VTNPSLMLVKSEVFGPIVIVSVKSDDEAIRANSTGYLQSAIFNDVNRALKLSRE 418
Qy 424 LEVGTGHNNKTRGPDNFPPLGLKGGAGVQIRYSIEAMTNVKSIVLDM 474
Db 419 LKFGAVINDSTRLWDSLPGGFKTGTIGREGVRETMLWENTENKLIATL 469

RESULT 13
D90323
hypothetical protein gapN-1 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
A:Accession: D90323
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: D90323
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-470 <KUR>
A:Cross-references: GB:AE006641; NID:gl3814871; PIDN:AAK41843.1; GSPDB:GN00155
C:Genetics:
A:Gene: gapN-1
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 31.3%; Score 750; DB 2; Length 470;
Best Local Similarity 36.6%; Pred. No. 4.5e-41;
Matches 171; Conservative 94; Mismatches 186; Indels 16; Gaps 7;

Qy 20 IELSPIDSSSLGFPVAMTREVVDHAKAGREALPAAALTVYERAYQLHKAADI 79
Db 7 IEVRSPSNLVIGTVKRMKDEVGEIEEAYKGPILSRMPLYKRTAILRKSEILEREQ 66

Qy 80 EETATVLAKESKAYNASVTVETADLIRYAAEE-GIRLSTADSGKKMA---STG 133
Db 67 ERLAKLAMEAGKPIRDSRVETLRASLFRHAAEEVGWLE---GKNYVDAYEYYPGNE 123

Qy 134 HKLAVIRQPGVILAIAPYVNLGSKIAPALIGNVNMFKPTQGSVGLVIAKAF 193
Db 124 NRIVSTREPIGVVTTALFPNFPINSFAHKVAPALAVGNSVVKPSINTPLAAIEMKIL 183

Qy 194 AEAGLPAGVFNITIGRSEIGDYIVVEHEVNFNFTGSPVQSGIKLA---GMRPIMLE 250
Db 184 VEAGLPDSAVRVVTVGYSSEIGDEIITHTPLVGLITLTGTQTGLKIAKASVLSGR- 242

Qy 251 LGKDGAGIVLADADLNAKQIVAGAYDSQRCATKRVLVVEEVADELAEKISENVAK 310
Db 243 LGGSDPIIILEDANIERASSIAVRARFEYAGQNCNAGKRIIVREIYDKFKAFNDKARA 302

Qy 311 LSVGDPDFNAT--VTPVIDDMSADFIESLVDARQKGAKE--LNEFKRDGRLLTPGLFDHV 367
Db 303 LKVGDPIDETITDVGFINKESENLSNVLDEAKVKGRVEILNKPESGSPFLTMVTP 362

Qy 368 TLMKLAWEPPGPIILPIIRVKDAEAAVANKSDFGLOSSVFTDFQKAFDIANKLEV 427
Db 363 SLDMVLKSEVFGPIVIVSVKSDDEAIRANSTGYLQSAIFNDVNRALKLSRELKFG 422

Qy 428 TVHNNKTRGPDNFPPLGLKGGAGVQIRYSIEAMTNVKSIVLDM 474
Db 423 AVIINDSTRLWDSLPGGFKSISIGREGVRETMLWENTENKLIATL 469

RESULT 14
H69230
NADP-dependent glyceraldehyde-3-phosphate dehydrogenase - Methanobacterium thermoautotro-
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
C:Accession: H69230
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; I

Db 123 PIDAHNGK-----GKLFYIRVPVIGIVSALTFFNPLNLSMKHVKAPALAAAGNAVILPK 176

Qy 179 PTQSSVGLVLAKAFAPAGVFNITGTGSGSIGIYVEHEEYVNFNTGSTPTGQRI 238

Db 177 SERTPLTLMLEIGLLEAGVPPKALSVIPGYG-DVGRAMTTPDVRVVSFGSRKVGDI 235

Qy 239 GKLWGRMPIMLELGGKADGIVLADADLNAKQIVAGAYDSYQRCQTAIKRVLVVEEAD 298

Db 236 ARQVKKVLELSEIGNSALIDHKGKLEKAVEKIVQGYALAGVCISVQRFVHEDLYD 295

Qy 299 ELAEKISENVAKLSVGPDPNAT-VTPVIDNSADFTIESLVVDARQKAK-ELNEFK--R 354

Db 296 EFKULKERVSKLVKGDPMNEDTDLGPMIAPSELERIQEWISEAVQKGAIEAGLRCAE 355

Qy 355 DGRLLTPELFDHVTLMKLAWEPPGPIIPRIYKDAEBAVAIAANKSPGLQSSVFPD 414

Db 356 DETFTSPTISLVPSDSKLFEEAFAPVAVVNFYKDIIEAIRWNSSDYLQGVFN 415

Qy 415 QKAFDIANKLEVCTVHINKTGRGPNFPFLGLKSGAGVQGIYRSIEAMTVKSVILDM 474

Db 416 KLAWKCIKEIAGGVLINEGPTFRADHMPYGVKYGKISGREGPKFAIEDYTIKTVIFDL 475

RESULT 17

T46684

A:Reference number: Z23132; MUID:20029269; PMID:10565539

A:Accession: T46684

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-491 <CRO>

A:Cross-references: ENML:U96338; NID:94808503; PIDN:AAA75634.2; PID:94808506

A:Experimental source: NCIMB 9866

C:Genetics:

A:Gene: pcha

A:Genome: plasmid pRA4000

C:Function:

A:Description: catalyzes oxidation of p-hydroxybenzaldehyde to p-hydroxybenzoic acid

A:Note: cofactor NADP+

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 30.8%; Score 736.5; DB 2; Length 491;

Best Local Similarity 35.8%; Pred. No. 3.6e-40;

Matches 171; Conservative 84; Mismatches 197; Indels 25; Gaps 7;

Qy 5 YQN----YVNGEWK--SSVNOITLSPIDSSLGFPVPMTRREVDHAKAGREALPAA 58

Db 8 YENWSLIQIAGWVRGKAGRLDVLDPFTQKLLQIPLANREDLDEAYRSARQAVAA 67

Qy 59 LTVYERAVLHKAADIIRKDEEATVLAKEISKAYNASVTEVVTADLIRYAAEGRIL 118

Db 68 CGPSERAAQMLNVRIFDERRSEIIDWIIRESG-----TRIKALIEWGAARATQ 118

Qy 119 ST----SADGGKMDASTGHKLAVIRROPVGIIVLAIPYVNLGSKIAPIALIGNVV 174

Db 119 ESASLPSRVHRIASDPVGSRYVREPLGVIIGISPNFPLHLTARSALAPALGNAC 178

Qy 175 MFKPTQSGVS--GLVLAKAFAPAGVFNITGTGSGSIGIYVEHEEYVNFNTGSTPT 233

Db 179 VIKPASDTPVTGGLLAHIFEEAGLPKGVLSVWVGSSEIGDAFVEHEVFPGISFTG 238

Qy 234 VQGRIGKLAG-----MRPIMLELGGKADGIVLADADLNAKQIVAGAYDSYQRCQTAIKR 289

Db 239 VGRNIGRIAGGEHLKHLVALELGNFPVVLADADLNAKQIVAGAYDSYQRCQTAIKR 298

Qy 290 VLVEEVADELAEKISENVAKLSVGPDPDNATVT-PVIDNSADFTIESLVVDARQKAKE 348

Db 299 IIVDSYDFEFVNRVAERVKSLYPGDPSTVTVGPVINAQLAGLQDIATASEGARV 358

Qy 349 LNEFKRDRLLTGLFDHVTLMKLAWEPPGPIIPRIYKDAEBAVAIAANKSPGLQSS 408

Db 359 WVEGAQGNVLPHPVADVADTADMEIAREEIEFGPLVGIQARDEAHALEANSSEYGLSSA 418

Qy 409 VTPTRDFOKAFDIANKLEVCTVHINKTGRGPNFPFLGLKSGAGVQGIYRSIEAMT 465

Db 419 VFTSLERGKVFARGIRAGMTINDIPVNDPNAPFGGKXNGLGRFNGDMALIEFT 475

RESULT 18

B83867

NADP-dependent glyceraldehyde-3-phosphate dehydrogenase BHI738 [imported] - Bacillus hal

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B83867

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: B83650; MUID:20512582; PMID:11058132

A:Accession: B83867

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-482 <STO>

A:Cross-references: GB:AP001513; GB:BA000004; NID:G10174345; PIDN:BA805457.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BHI738

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 30.7%; Score 735; DB 2; Length 482;

Best Local Similarity 35.0%; Pred. No. 4.4e-40;

Matches 164; Conservative 90; Mismatches 209; Indels 6; Gaps 3;

Qy 9 VNGEKSSVNOITLSPIDSSLGFPVPMTRREVDHAKAGREALPAAALTYERAQYL 68

Db 12 ISDQWQGGQSSIEVDGTNTHVIAIVPAATEAKAIAEAKEGEIAIARQMPITHEAAIL 71

Qy 69 HKAADIIRKDEEATVLAKEISKAYNASVTEVVTADLIRYAAEGRILSTADEGGK 128

Db 72 KRTAEIVTDRLEFAQTIKESKTIARESEVRCIETLTLSGEAKRLHGETIPFSQM 131

Qy 129 DASTGH--KLAVIRROPVGIIVLAIPYVNLGSKIAPIALIGNVVMPKPTQSGVSG 186

Db 132 ---PGHRRVGYFRFPPIGIIVAITPFNDPLNVAHKIGIAPAGNSLIKPSFTFLSA 188

Qy 187 LVLAKAFAPAGVFNITGTGSGSIGIYVEHEEYVNFNTGSTPTVQRIKGLAGMP 246

Db 189 LRLVKALEAGLPKKIVQVITGHSGVIGFTLTKHRDVLISFTGGYETGEKIARSAGVK 248

Qy 247 IMLELGGKADGIVLADADLNAKQIVAGAYDSYQRCQTAIKRVLVVEEVADELAEKISE 306

Db 249 LAMELGSPTVILQDAELMEAVASCVSAFCAGQNCIGVQRIYVEQSVFNSFISFVA 308

Qy 307 NVAKLSVGPDPDNAT-VTPVIDNSADFTIESLVVDARQKAKELNEFKRDLITPGLFD 365

Db 309 QTKQLRGLGKQSEETDIPMISEKAKRIERWEEAKEGARVLTGGRRTGAYFEFTVLT 368

Qy 366 HTVLMKLAWEPPGPIIPRIYKDAEBAVAIAANKSPGLQSSVFPDFOKAFDIANKLE 425

Db 369 NVSPSLRAKEAFAPVILIEGVHSTEARANDVDVGLQAGLFTNNLTAAFSATEKLE 428

Qy 426 VGTVHINKTGRGPNFPFLGLKSGAGVQGIYRSIEAMTVKSVILDM 474

Db 429 VGGIWNDSVDRIDAMPFGIKKSLGREGVYIAEEMTEQKVAFHL 477

RESULT 19

F65045

succinate-semialdehyde dehydrogenase [NAD(P)] (EC 1.2.1.16) - Escherichia coli (strain K-

C:Species: Escherichia coli

[illegible]

Db 427 FRVGEALYGVGIN--TGLISNEVAPFGGKASGLGREGSKYGLDYLIKWCIGL 482

RESULT 24
D85913
succinate-semialdehyde dehydrogenase [imported] - Escherichia coli (strain O157:H7, subs
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C:Accession: D85913
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85913
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE005174; MID:g12517094; PIDN:AAG57768.1; GSPDB:GN00145; UWGP:239
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: gabD
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 30.0%; Score 717; DB 2; Length 482;
Best Local Similarity 33.9%; Pred. No. 6.4e-39;
Matches 162; Conservative 92; Mismatches 208; Indels 16; Gaps 7;

Qy 6 QNTVNGEMKSSVN--QIEILSPIDSSLSGFVPAMTREETVEVDHAKAGREALPAAALTYVVE 63
Db 12 QALINGELDANNGEVIDVTPNANGKLSGVPKMGADETRAIDAANALPWRALTAKE 71
Qy 64 RAQYLKAADIIRKKEELATVLAKEISKAYNASVTEVVTADLIRYAABEGIRLSTSD 123
Db 72 RANILRNWFLMWEHQDRLMTLEQGGPLAEAKGEISYAASFIEWAEGRKGYDGTI 131
Qy 124 EGGMDASTGKLIARVIRPQGVILAIAPYVNVNLSGSKIAPALIGNVNMFKPPTQGS 183
Db 132 PGHQAD-----KRLIVKQPIGVTAITPWFNPAMITRAGPALAACTWLPKASQTP 186
Qy 184 VSGIIVAKAFABAGLPAGVNTITGRSGEIGDIYVEHEEVNFINTGTPVQGRIGKLAG 243
Db 187 FSALAAELAIAGIPAGVENVVTGAGAVGNELTSNPLVKLSPTGSTEIGRLMEQCA 246
Qy 244 --MRPIMLEGGKDGAGIVLADADLNAAKQIVAGAYDSQRCCTAKKVLVVEEVADELA 301
Db 247 KDIIKVSLEIGGNAPFIVFDADLDKAVEGALASKFRNAGQTCVCANRLYVQDGVYDFA 306
Qy 302 EKISENVAKLSVGDPFD-NATVTPVIDDNSADFIESLVVDARQKAGEL---NEFKRQGR 357
Db 307 EKLQAVSKLHIGDGLGKVTIGPLIDEKAVAKVEHIALEKGRVVCCKAHERGNN 366
Qy 358 LITPLGPDVHTLDMKLAWEEPFGPIILIRVKDAEEAVAIANKSDFGLOSSVFTDRPQKA 417
Db 367 FQQTILVDPANAKVSEETFGPLAPFRFDEADVIAQANDTESGLAAFYFARDLSRV 426
Qy 418 FDIANKLEVTHINKTG-RGPDNPFPLGLKSGAGVQGIYSIEAMNKSIVLDM 474
Db 427 FRVGEALYGVGIN--TGLISNEVAPFGGKASGLGREGSKYGLDYLIKWCIGL 482

RESULT 25
C84250
glyceraldehyde-3-phosphate dehydrogenase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: C84250
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Letthausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84250
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-480 <STO>
A:Cross-references: GB:AE004437; MID:g10580497; PIDN:AAG19367.1; GSPDB:GN00138
C:Genetics:
A:Gene: gap
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 29.8%; Score 713.5; DB 2; Length 480;
Best Local Similarity 35.0%; Pred. No. 1.1e-38;
Matches 165; Conservative 86; Mismatches 210; Indels 11; Gaps 6;

Qy 8 YVNGEMKSSVNOIEILSPIDSSLSGFVPAMTREETVEVDHAKAGREALPAAALTYVVEAQY 67
Db 13 FVAGEWVSDEGFLPVSDLDGDTFASVAADTSDAERALSAAATGVHADLRRTTVPERVIEW 72
Qy 68 LHKAAADIIRKKEELATVLAKEISKAYNASVTEVVTADLIRYAABEGIRLSTSDGGK 127
Db 73 LESTADGIRREDELAEVIVREACKPISSARGEVQSAERFDRVAGE-LRHLTGTYTG- 130
Qy 128 MDASTGHK--LAVIRPQGVILAIAPYVNVNLSGSKIAPALIGNVNMFKPPTQGSVS 185
Db 131 --TTAGHEDWQAIKVHEPMGTGLITPYNPLSTWALQVAPALAAAGNAVIVKPAKTPIS 188
Qy 186 GLVLAKEAFABAGLPAGVNTITGRSGEIGDIYVEHEEVNFINTGTPVQGRIGKLAGMR 245
Db 189 GAILADTAADAGLPDGAANFVPGSESVIGDPLASDARVDALAMTSGSGAGEHVARQSGIT 248
Qy 246 PIMLEGGKDGAGIVLADADLNAAKQIVAGAYDSQRCCTAKKVLVVEEVADELAEKIS 305
Db 249 RLEMELGNAPALVPEADADLDAADAATAAGSLKYGQRCQSAVSRVLAHESVHDELVGRID 308
Qy 306 ENVAKLSVGDPFD-NATVTPVIDDNSADFIESLVVDARQKAGELNEFKR---DG-RLLT 360
Db 309 DAMAEWSIGLDFDTDTTGLPLVVSADQADWVAELVDVADVRGATVVRGGEHVEDGVHYE 368
Qy 361 POLGPDVHTLDMKLAWEEPFGPIILIRVKDAEEAVAIANKSDFGLOSSVFTDRPQKAPDI 420
Db 369 FTLLADVPDARIVDEQFGVCATVTTDEDDAVRTANGSELALDAAVFTADHDRAMRV 428
Qy 421 ANKLEVTHINKTG-RGPDNPFPLGLKSGAGVQGIYSIEAMNKSIVL 472
Db 429 ASRVNAGAVRINGAPSHGLGVDPFGNDASGIGREGLDSTIHEFVREKSIIL 480

RESULT 26
T44939
glyceraldehyde-3-phosphate dehydrogenase (NADP) (EC 1.2.1.9) [validated] - Thermoproteus
C:Species: Thermoproteus tenax
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Jun-2002
C:Accession: T44939
R:Brunner, N.; Brinkmann, H.; Siebers, B.; Hensel, R.
J. Biol. Chem. 273, 6149-6156, 1998
A:Title: NAD+-dependent GAPDH from Thermoproteus tenax - the first identified archaeal me
A:Reference number: Z22875; MUID:98165789; PMID:9497334
A:Accession: T44939
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-501 <BRU>
A:Cross-references: EMBL:Y10625; PIDN:CAA71651.1
A:Experimental source: strain KRA1
C:Genetics:
A:Gene: GAPN
C:Function:
A:Description: catalyzes the NAD+-dependent, phosphate-independent irreversible oxidation
A:Note: Thermal stability: 30% residual activity after 100 min at 100 degrees C
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: Oxidoreductase

Query Match 29.8%; Score 713; DB 2; Length 501;

126 GKMDASTGHKLIIVRRQPVGVILVLAIPYVPNLGSGKIAPALIGWVMKEPPTQGVS 185
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
130 DHKQSVEVGKENIIVREPOGWGGIIISPNFPLNISRAVAPAAGAASVLVKPASNTFTVT 189
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
186 -GLVTLAKFAEAGLPAEGVNITIKRGEIDGYIVHEEVNFINTGSTPVQRIGKLGM 244
 ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
190 GLLIAKFVAAGLPFGVNVNVTKGSEICDRJAGHPHADVISFTSGTAIRQVAATAGE 249
 ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
245 RPIV--LELGKKDAGIVLADDLNAKQTVAGADYGQRCATAIKRVLVVEEVADELAS 302
 -::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
250 NLALPAMELGGNNVHIVTEADLDVVDSVGFSLHGSGCISINRHVVHEDVDVEYVE 309
 :-:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
303 KISENVALKSVDGPDFDNATV--PVIDNSADTFESLVWDARQKGAKELFNEFKRGRLLTP 361
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
310 RLTERAAALPGTCAHDADTVWGPIIDSORDQLDIERTVDAGATLETCTGDHDDGLVPEP 369
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
362 GLFDHTVDMKLANEBPFEPFLPIIRVKDAEEAATANKSDFGLQSSVFTRDFQVAFDIA 421
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
370 TVLSDATNDMAAACNEHFGPVAVIPFSDDDEAVEIANSTEYGLSGSVHAGDTGRARDIA 429
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
422 NKLEVGTTHINKTRGPDNFFPLGLKSGAGVGQIRYSIEAMTN 466
 :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
430 DRIDTGMIHINDQPWQEHLPPFGYKNKSLG---RYNGEIVLN 470

RESULT 28

Ald277
aldehyde dehydrogenase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD2277
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2277
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <CUR>
A:Cross-references: GB:BA000019; PIDN:BAB75470.1; PID:g17132905; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3771
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 29.7%; Score 710.5; DB 2; Length 498;
Best Local Similarity 33.0%; Pred.No.1.8e-38;
Matches 159; Conservative 105; Mismatches 199; Indels 19; Gaps 8;

QY 6 QNYVNGEMKSNVQBIILS--PIDDSL-GFPAMTREVEDHAMKAGREALPAWAALTYY 62
DB 8 QNYINGQWLNAETTTILSHNPADKSEIVATEPRSQADDVRAVAABAQAAYGSWRKVAP 67

QY 63 ERAQYLHKADIITRKKEIATVLAKEISKANVASVTEVVRTADLRVAAEGRLSTSA 122
DB ||||: :: : |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
QY 68 AAAYIFYPVCELLHQEKELAQGISRMGRPTEARGDVQGVDCAFVSAGERL F--- 124
DB ||||: :: : |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

QY 123 DEGRKMDSATHCLVIIRQPVGVILAIPYVPNLGSGKIAPALIGNVVMKPPTOG 182
DB |--GOTTPEMENEFMTMEPIGVCALLTPWFVPAIPCWKAMPALVCNTVILKPAEDT 182

QY 183 SVSGLVLAKAPAFAGLPAGVFNITTKRGSEIGDYIVEHEEVNFINTGSTPVQRIGKLA 242
DB ||||: :: : |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
QY 183 PACATKLJEIFAAGLPFGVINLVHGVGEEVKALVEHPNIIDLVSFTGSSATGAYGETC 242
DB ||||: :: : |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

QY 243 GM--RPIMLEGKDAGIVLADLDNAKQTIVAGADYQRCATAIKRVLVVEEVADEL 300
DB |-:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
QY 243 GRTHKRVLCLEMGGNQCVNEDADELALDGALWGAFGTGTCATSRILHRDIKEKF 302
DB ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

QY 301 AEKISENVALKSVDGPDFDNAT-VTFVIDNDSADFIESLVWDARQKGAKE L--NEFKDGR 357

```

303 TTWLRRTSQRLEAGTETPTDGIPIINRQLOQRVHEVYNNIAREBGAKILIGGEIATEGQ 362
358 L-----LTPGLFDHVTLDKMLAWBPFGPILPIIRVKDAEEAVATANKSDFGLQSSVFR 412
363 LOQGYFPQFTILDVNTQMRVAREETFGPVVALIEVSTFEAAIALNDTKYGLSSVYTR 422
413 DFOKAPDIANKLVGTVHNNKTRGPNDFPLGLKSGAG-VQGIYRSYEAMTNVKSIV 471
423 DINRAVAMRDIEAGITVINGPTGAHVLPFGVKGTGNHREAGTALDVTFEKSYI 482
472 LD 473
483 VD 484

RESULT 29
C83774
succinate-semialdehyde dehydrogenase BH0995 [imported] - Bacillus halodurans (strain C-1
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C83774
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C83774
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-468 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04714.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 29.1%; Score 697.5; DB 2; Length 468;
Best Local Similarity 33.4%; Pred. No. 1.1e-37;
Matches 158; Conservative 93; Mismatches 209; Indels 13; Gaps 5;

QY 8 YVNGEKSSVNOQIEILSPIDSSLGFPVPMTEEVVDHAKGREALPAWALTIVYRAQY 67
DB 4 YINGENWRSCKTLDVTNATGEVDTVPAGKKEAEALVAAGAAVEAPPQWSSQTASERSY 63
QY 68 LHKAAIDIERDEKEIATVLAKEISKAYNASVTEVVRADIRYAAEEGRLSTSDSGGK 127
DB 64 LMRWFQIDQDQDEIGTEINTKEQKPLREAGVEQYANSFTQWYAEAKRI-----YGD 118
QY 128 MDASTGHKLAVIRROPVGIIVLAIPYNNPVNLSGKIAPALIGGNVVMFKPPTQGSVGL 187
DB 119 IPASAINKRLVQKQPVGVIAAITPNFPAAMITRKVAPALAAAGCTAIVKPAEQTLPTAL 178
QY 188 VLAKAPAEAGLPAGVNTITGRSGEIGDYIVHEHEVNFINTGSTVPGORI--GKLAMR 245
DB 179 KLAQAEAEAGIPAGVLNVITNAQDICEAWLEDSEVRKRTITFTGSTVGVKLMRGAATVK 238
QY 245 PIMLEGGKAGIVLADADLNAKQIVAGDYDSORCTAIKVLVVEEVADELAEKIS 305
DB 239 KISLELGHAPFIMDDANLEAVDQVIASKFRNAGQTCVCANRIVAREIAEAFTEKFA 298
QY 306 ENVAKLSVGDPPF-DNATVTPVIDONSADFTESLVVDARQKGAKEINFRDGLRILTEGLF 364
DB 299 AKVWELKVGNGLEEGVTIGLIDKAAVEKVEAHIDALKKGQVTVGGRWTHFPEPTII 358
QY 365 DHVTLDMKLAWBPFGPILPIIRVKDAEEAVATANKSDFGLQSSVFRDFOKAFDIANKL 424
DB 359 TGAATDEMLCMEETFGPLAPVATEDTEEEVIERANHTPYGLAAYVFTENIGRAIRLSEKL 418
QY 425 EYGVTHNNKTRGPD--NFPLGLKSGAGVQGIYRSYEAMTNVKSIVLDMK 475
DB 419 EYGIYGVNDGW---PSVAQPPGGKSGSLGREGGKYGIEEYLEVYKVSVAFK 468
```

```

RESULT 30
B83901
aldehyde dehydrogenase BH2010 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B83901
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B83901
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-485 <STO>
A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05729.1; GSPDB:GN001
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 28.5%; Score 682.5; DB 2; Length 485;
Best Local Similarity 33.4%; Pred. No. 1.1e-36;
Matches 159; Conservative 96; Mismatches 210; Indels 11; Gaps 5;

QY 6 QNYVNGEW--KSSVNOQIEILSPIDSSLGFPVPMTEEVVDHAKGREALPAWALTIVYE 63
DB 7 ROFLAGEWIEGASTQITITNPNYPTGKIEHKIGASLDDLDQAVRAAKNAKWEQTLPAQ 66
QY 64 RAOYIHKAAIDIERDEKEIATVLAKEISKAYNASVTEVVRADIRYAAEEGRLSTSD 123
DB 67 KOEVLKAAALAEKDEVERLVNAGSSIYKASIEWGATLQTIKVAATFFLRM----- 121
QY 124 EGGKMDASTGHKLAVIRROPVGIIVLAIPYNNPVNLSGKIAPALIGGNVVMFKPPTQGS 183
DB 122 ESKILPSNIGKENRIYRSAGKGVISPEPFLVAMRSVAPALATGNAVVLKCSSDAP 181
QY 184 V-SGLVLAFAEAGLPAGVNTITGRSGEIGDYIVHEHEVNFINTGSTVPGQRIQKLA 242
DB 182 ITSGLLIAELFEAGLPGLVNVVVGKAAEIGDAFVTHPIPKLISFTSGSTEVEGRHIAQL 241
QY 243 G--MRPIMLEGGKAGIVLADADLNAKQIVAGDYDSORCTAIKRVLVVEEVADEL 300
DB 242 ABELKETALELGGNNVMIVDDADIEKAEKAAVGVKFLHQGQICMALNRIIVDASIDYSP 301
QY 301 ABKISENVAKLSVGDPPDNAT--VTPVIDDMSADFTESLVVDARQKGAKEINFRDGLRL 359
DB 302 VEVFKEKVSQLOQTGNPAEPATLIGPLINYKQIGRIQLVKESVAQGVKVLGHHVQGNLM 361
QY 360 TPLGFDHVTLDKMLAWBPFGPILPIIRVKDAEEAVATANKSDFGLQSSVFRDFOKAFD 419
DB 362 SPTILSEVTNDMPVAKKEITFGPIAPIIKAKDEAEATAINDSPYGLSGSIFTGSLHRGVQ 421
QY 420 IANKLEVTVHNNKTRGPDNDFPLGLKSGAGVQGIYRSYEAMTNVKSIVLDMK 475
DB 422 VAKQIDTGNLHVNDQPVNEEAHISFGGEKSGIGRFGGEWLDKFTVTKWISIQEK 477

RESULT 31
AE0839
succinate-semialdehyde dehydrogenase [NAD(P)] [EC 1.2.1.16] - Salmonella enterica subsp.
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0839
R:Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moulle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0839
A:Status: preliminary
```


A:Molecule type: DNA
 A:Residues: 1-482 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD05900.1; PID:916503874; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY2911
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 28.5%; Score 682; DB 2; Length 482;
 Best Local Similarity 33.1%; Pred. No. 1.2e-36;
 Matches 158; Conservative 95; Mismatches 203; Indels 16; Gaps 7;

QY 6 QNYVNGWKS--VNOIETLSPIDSSLGFPVPMATREEDVHAMKAGREALPAWALTVE 63
 DB 12 QAFIDGWRDARGDVIQVSPNPKPLGNVPMGAETRDAINANRALPAWALTVE 71
 QY 64 RAQYLHKAADIIRDEKEIATVLAKETSKAYNASVTEVTRPADLIRYAAEGIRLSTAD 123
 DB 72 RANILRRWENLMMHQDDLRLMTLEQKPLAEAKGEISYAASFIEWFAEGKRIYGTI 131
 QY 124 EGGKMDASTGHKLAVIRPQVIGVIAIAPYNPVNLGSKIAPALIGNNVMEKPTQGS 183
 DB 132 PGHQT-----KELLVQPIGVTAATPWNFPESAMITKAGPALAAGCTVWIKFASQTP 186
 QY 184 VGLVLAKAFABAGLPAGVNTITGRGSEIGDYIVHEEVNFNFTGSTPVQRIKLAG 243
 DB 187 FSAIALBELAQRIAGLVNFVVTGSAGDIGGELSNPLVKLSFTGTEIGRQLMEQCA 246
 QY 244 --MRPIMLELGGKAGIVLADLDNAAKQIVAGDYSGORCTAIKRVLVVEVADLA 301
 DB 247 KDIKVSELGGNAPFIVDDADLDKAVEGALASKPRNAGQTCVCANRLYVQDGYDRFA 306
 QY 302 EKISENVAKLSVGGPDP-NATVTPVIDNSADFIESLVVDAROKAKEL---NEFKRGR 357
 DB 307 EKLQAVNKLAVGGGQADVAITGGLIDEKAVKVOBHTADALEKRGAVITGGEAHLGNN 366
 QY 358 LITPGFLPHVTLDMKLAWPEPPGPIPIRVKDAEVAIAANKSDFGLSSVFTDRDFOKA 417
 DB 367 FQPIILADVDPAKAKETGPIPLFRFSEADVIQANDTEFGLAAYFVARDLSRV 426
 QY 418 FDIANKLVGVTHNNKTG-RGPDNFFPLGKSGAGVQGIYRISEAMTNVKSIVLDM 474
 DB 427 FRVGEALYGVIGIN--TGIISNEVAPFGKISGLGREGSKYGIEDYLEIKYMCIGL 482

RESULT 32
 T44987
 C:Species: Haloflex volcanii
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
 C:Accession: T44987
 A:Description: Hereditary instability of the megaplasmid pHV3, and filamentation in the
 A:Reference number: Z22886
 A:Accession: T44987
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-496 <PAR>
 A:Cross-references: EMBL:U95374; PIDN:AAB71806.1
 A:Experimental source: strain DS2
 C:Genetics:
 A:Map position: megaplasmid pHV3
 A:Genome: plasmid
 A:Note: expressed during exponential growth
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 28.5%; Score 682; DB 2; Length 496;
 Best Local Similarity 34.6%; Pred. No. 1.2e-36;
 Matches 166; Conservative 85; Mismatches 207; Indels 22; Gaps 8;

QY 8 YVNGWKS--QIEILSPIDSSLGFPVPMATREEDVHAMKAGREALPAWALTVE 65

DB 15 YIDGWTESGESIAVEDPSTREITVAHVPRGTEDVDAAYEAAEAQSWAEPARQ 74
 QY 66 QYLHKAADIIRDEKEIATVLAKETSKAYNASVTEVTRPADLIRYAAEGIRLSTAD 125
 DB 75 EYVEQFQALNEVEDEIDLLAHVGGSRIMGETSIQIASHASEATLPRM-----RG 129
 QY 126 GKMDASTGHKLAVIRPQVIGVIAIAPYNPVNLGSKIAPALIGNNVMEKPTQGS 185
 DB 130 EHVASNPGRKENVQKNPKGVTVISPNFPLNSMRAVAPAAVAGNAVVLKPFSTNSPIT 189
 QY 186 -GLVIAKAFABAGLPAGVNTITGRGSEIGDYIVHEEVNFNFTGSTPVQRIKLAGM 244
 DB 190 GELLFAKLFEETDLPEGVVNVVTVGRGSEIGDRVAGHPESDVVAFSTGVGRVSGIAGE 249
 QY 245 R--PIMLELGGKAGIVLADLDNAAKQIVAGDYSGORCTAIKRVLVVEVADLA 301
 DB 250 NIAVAM-ELGNNAHVVTGGADVDRADAAVFGSVFGQVCISINRHIVHESVYDEVV 308
 QY 302 EKISENVAKLSVGGPDPNATVT-PPVIDNSADFIESLVVDAROKA-----KELNEF 352
 DB 309 EKITERAAELFVGSAREDDTVVGPIDESQORDEMLGYVEETVAAGATLETGGSTADLDGV 368
 QY 353 KRPGRLITPGFLPHVTLDMKLAWPEPPGPIPIRVKDAEVAIAANKSDFGLSSVFT 412
 DB 369 D-DSLWQVFTVLSGVTNDMAAARNEHFGPIAPVPSVDVDEAVELANATYGLSGSVHAG 427
 QY 413 DFQKAFDITANKLVGVTHNNKTGPDNFFPLGKSGAGVQGIYRISEAMTNVKSIVL 472
 DB 428 DIAATKEIALNETGVNVHNDPINDENHVPFGTSGASGVGTNSDAFLDEITDKWISL 487

RESULT 33
 C69584
 aldehyde dehydrogenase aldy - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: C69584
 A:Authors: Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteri
 R.; Kunst, F.; Ogawara, N.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 C.; Bron, S.; Brouillet, S.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei,
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: C69584
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-485 <KUN>
 A:Cross-references: GB:Z95123; GB:AL009126; NID:G2636240; PIDN:CAB15909.1; PID:G2636418
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: aldy
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 28.1%; Score 673; DB 2; Length 485;
 Best Local Similarity 33.9%; Pred. No. 4.6e-36;
 Matches 161; Conservative 95; Mismatches 201; Indels 18; Gaps 6;

QY 6 QNYVNGWKS--SSVNOIEILSPIDSSLGFPVPMATREEDVHAMKAGREALPAWALTVE 63
 DB 8 KSPFNGKWTGSGRTEDILNPVDQSVITATSLATGQLEDAFIAQAKQAKWAKSTTD 67
 QY 64 RAQYLHKAADIIRDEKEIATVLAKETSKAYNASVTEVTRPADLIRYAAEGIRLSTAD 119

Db 68 RKAVLQKARGYLHNRDDIIMIAETGTTIIKSTIELEOTIAILDAMTYTGGVGKE 127
 QY 120 TSADEGKMDASTGHKLAVTRPOGIVLAPYVNVNLSGSKIAPALICGNVWFKPP 179
 Db 128 VPSDIEGKTNK-----IYRLPLGVITSSISPFNMLSMRSIAPALAGNSVVRKPD 179
 QY 180 TQGSVS-GLAVAKAFAPAGAGPAGVFNITGRSGEIGYIVIEHEEVNFNFTGSPVQORI 238
 Db 180 IQTAISGGIIIAKAFHAGLPGVNLVMTDVKEIGDMITNPILFISFTGTAAGRHI 239
 QY 239 GKLAG--MRPIMLEGGKDGAGIVLADADLNAAKQIVAGAYDSGQRCTAIRVLVVEV 296
 Db 240 GEIAGRAFKMALELGGNNPFAVLSDADVDAADAAIFGKFIHQGICIMINRIIVHQDV 299
 QY 297 ADELAEKIENVAKLSDGDFPDNATVT--PVIDNSADFIESLVVDARQKAKELNEFKRD 355
 Db 300 YDFEVEKFTARVQKLPYGDQDTPKTVGVGELINERQIEKALEIEQAQTDGIELAVEGKV 359
 QY 356 GRLLTGLDFHVTLDKLAWEPPGPIPLIRVVDAAEVAIAANKSDFGLOSSVFTDRFO 415
 Db 360 GNVLPVVFVGDNNSKIAQTELFAPATIIKAGSDQEAIDMANDETVGLSSAVFTSDLE 419
 QY 416 KAFDIANKLEVGTTHNNKTRGPDNFPFELGLKSGAGVQGIYRSIETAMTNVKS I 470
 Db 420 KGEKFAQLQDSGTHVNDQSVNDSPNIAFGNKASGVGRFGNPVVEEFTVTWKI 474

RESULT 34
 A87547
 vanillin dehydrogenase [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C/Accession: A87547
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of Caulobacter crescentus.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: A87547
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-470 <STO>
 A/Cross-references: GB:AE005673; NID:gl3423939; PIDN:AAK24373.1; GSPDB:GN00148
 C/Genetics:
 C/Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 28.0%; Score 669.5; DB 2; Length 470;
 Best Local Similarity 34.9%; Pred. No. 7.4e-36;
 Matches 155; Conservative 88; Mismatches 190; Indels 11; Gaps 4;

QY 36 AMTREVVDHAKAGREALPAWAAVTYVRAQVYLHKAADIIRKKEIATVLAKEISKAYN 95
 Db 24 AMTADQARVADAAQAALPTWSALGPNARRALLSKAAALEARADDFVAAAMGIGATEG 83
 QY 96 ASVTEVVRTADLIRYAEIEGLSTSADEGKMDASTGHKLAVTRPOGIVLAPYVNVN 155
 Db 84 WARFNMLAASVYREAA-----ALTTQVSGEIVPSDKPGCLKAVMAVREPVGVLGIAPWNA 138
 QY 156 PVNLSGSKIAPALICGNVWFKPPQTQSGVSLVAKAFAPAGVFNITGR---GSE 212
 Db 139 PIIIGVAVATPLACNTVWLKASECPRTHIELIAEFAAGLPEGALSIVTNAPADAAE 198
 QY 213 IGDYIIVEHEEVNFNFTGSPVQGIKLAG--MRPIMLEGGKDGAGIVLADADLNAAK 270
 Db 199 VVGALIDHPAVRRINTGSTAVGKIIAKRAAEHLKPVLELGGKAPLIIVLEDAIDAEVK 258
 QY 271 QIVAGAYDSGQRCTAIRVLVVEEVADELAEKIENVAKLSDGDFPDNAT--VTPVIDDN 329
 Db 259 AAAGFAPMNGQCMSTERRIIVDAIADAFVAKFAAKASSIAGVDPREGKTPGLGAVDLK 318

QY 330 SADFIESLVVDARQKAKELNEFKRDGRLLTGLDFHVTLDKLAWEPPGPIPLIRVK 389
 Db 319 TVTHVQGLVADALADGAVSGGPGANGVLMPATVVDKVTAPMLFREESFGVAVIRAR 378
 QY 390 DAEBAVAIAANKSDFGLOSSVFTDRFOKAFDIANKLEVGTTHNNKTRGPDNFPFELGK 449
 Db 379 DEEHAIALANDTEYGLSASVFTDRFADRLKLVARRIQSGICHVNGPTVHDEAQMPPGGVKA 438
 QY 450 SGAGVQGIYRSIETAMTNVKSIVLD 473
 Db 439 SCYGRFGKAGVDAFTELRWITVE 462

RESULT 35
 D83613
 succinate-semialdehyde dehydrogenase PA0265 [imported] - Pseudomonas aeruginosa (strain I
 C:Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: D83613
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brj
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lam,
 J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10984043
 A/Accession: D83613
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-483 <STO>
 A/Cross-references: GB:AE004464; GB:AE004091; NID:g9946099; PIDN:AAG03654.1; GSPDB:GN0011
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: gabD; PA0265
 C/Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 27.9%; Score 668; DB 2; Length 483;
 Best Local Similarity 33.2%; Pred. No. 9.7e-36;
 Matches 158; Conservative 90; Mismatches 212; Indels 16; Gaps 7;

QY 6 QNYVNGEWSVN--QIETLSFIDSSIGFVPAMTREVVDHAKAGREALPAWAAVTVE 63
 Db 12 QAYVDGAWVDADNGQTIKVNNPATGEIIGSVPMGAAETRRRAIEADKALPAWRAITAKE 71
 QY 64 RAQYLHKAADIIRKKEIATVLAKEISKAYNASVTEVVRTADLIRYAEIEGLSTSD 123
 Db 72 RANKLRWFDLMTENQDLDLARLMTIEQKPLAEAKGEIAYAAASFEWFGSEAKRIYGTI 131
 QY 124 EGGKMDASTGHKLAVTRPOGIVLAPYVNVNLSGSKIAPALICGNVWFKPQTQGS 183
 Db 132 PGHQPD-----KRIIVIKQIPIGVTAAITPNWFFSAMITRKAGPALAAGCTMWLKPSATP 186
 QY 184 VGLVLAKAFAPAGVFNITGRSGEIGDYIIVEHEEVNFNFTGSPVQGIK--L 241
 Db 187 YSALALAEALAEARAGIKPGVFSVWTGSGAGEVGGELTSNPVLEKLTFTSGTEIGRLMAECA 246
 QY 242 AGRRPIMLEGGKDGAGIVLADADLNAAKQIVAGAYDSGQRCTAIRVLVVEEVADELA 301
 Db 247 QDIKKVSELGGNAPFIVDDADLDAAVEGALISKYRNNGQTCVCANRLYVDQGVYDAFV 306
 QY 302 EKISENVAKLSDGDFPD--NATVTPVIDNSADFIESLVVDARQKAKELNEFK---RGR 357
 Db 307 DKLKAVALKINIGLEAGVTGPELIDAKAVKVEEHTADAVSKGKAVSGGKPKHALGCT 366
 QY 358 LLTGLDFHVTLDKLAWEPPGPIPLIRVVDAAEVAIAANKSDFGLOSSVFTDRFOKA 417
 Db 367 PFEPIILVDPKVALVSKDEFEGPLAPVFPFKDEAEVIAVMSNDTEFGIASYFYVADLARV 426
 QY 418 FDIANKLEVGTTHNNKTR--GPDNFPFELGLKSGAGVQGIYRSIETAMTNVKSIVL 472
 Db 427 FRVAGQLEYGWVGIN--TGLISNEVAPFGGKIKASGLGREGSKYGIEDYLEIKYLCL 480

RESULT 36

[illegible]

Query Match	27.7%	Score 664;	DB 2;	Length 477;
Best Local Similarity	35.3%	Pred. No. 1.7e-35;		
Matches 170: Conservative	80:	Mismatches 603:	Indels 30:	Gaps 12

Qy 118 LSTSADE-----GGKMDASTGHKLAVIRQPGIVLAIAPNYPVNLGSKIAPALIGG 171

Db 133 ABLDVKETRGDYVGGWSSETLEAVVRKEPVGVVLVSIPEVPLFDVTKVIVTVTVIG 192
 QY 172 NVWFKPPTGSGVSLVLAFAAGLPAAGVNTITGRSGLGDIYVEHEEVNFINTGS 231
 Db 193 NAIILKPPSTPLPILMLAKWELASFPKDSFAITIFGRDM-NKVVGDKRIQALISUTGS 251
 QY 232 TPVGRTGKLAGMRPIMLELGGKAGIIVLADADLONAAKQIVAGAYDSGQRCTAIVKVL 291
 Db 252 TETGEVVRNAGIKQIFELGGDPAIVLSDADLAWAQRJAAGILISYTGQRCDSEKVL 311
 QY 292 VVEEVADELAEKISENAVAK-LSVGDPPDN-ATVTPVIDNSADPIESLVVDARQKAKEL 349
 Db 312 VEEVYDTLKDILIKELTKSVKVGDPDPLTIVGVIDTKVDEWEKAIDKAVEKGGKIL 371
 QY 350 NEFFKSDG-RLITGLFD---HVLDMKLAWEPFGPILPIIRVKDAEBAVAIAKNSDGL 405
 Db 372 FGKRLGPTTIEVLEAPKETLKDMDYFNKEVFAFASALLIKVKNIDEALEISNRKYL 431
 QY 406 QSSVFTDFOKAFDIANKLEVTGTHINNKTRGPDNFFFLGKSGAGVQGIKXISBAMT 465
 Db 432 DAAIFGDKIRKQRFLEVGAIYINDYPRHGIGVYFPGKDSGIGREGIGYTIQVIT 491
 QY 466 NVKSVILDMK 475
 Db 492 AYKSVIVNYK 501
 RESULT 41
 E96825
 Hypothetical protein T8K14.14 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: E96825
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ausen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: E96825
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-509 <STO>
 A:Cross-references: GB:AE005173; NID:G4835765; PIDN:AAD30232.1; GSPDB:GN00141
 A:Gene: T8K14.14
 A:Map position: 1
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 Query Match 27.6%; Score 661.5; DB 2; Length 509;
 Best local similarity 33.4%; Pred. No. 2.7e-35;
 Matches 165; Conservative 90; Mismatches 208; Indels 31; Gaps 8;
 QY 6 QNYVNGEWSVNVQ--TEILSPIDDSISLGFVPAMTVEEDVHAKGREALPAWALTVE 63
 Db 19 QGLIGKWLDSYDNKTIKVNPTATGEIADVACMGTKETNDATASSYEAFTSKSLTAGE 78
 QY 64 RAQLHLKAAIDIERD-----KEEIAVLAKIEISKAYNASVTEVVRTADL 107
 Db 79 RSKVLRRVNLRSBPLHFTKHYDILLIAKHELQQLITLQKPLKEAIGVAYGASF 138
 QY 108 IRYAABEGIRLSTSADEGKMADSTGHKLAVIRRPQVGIIVAIAPYVFNLSGSKIAPAL 167
 Db 139 IEYAAEAKRY-----YGDIIIPNLSDRLLVLKQPVGVGAIITPMNFFLAMITRKVGPA 193
 QY 168 LIGGNVMPKPTGSGVSLVLAFAAGLPAAGVNTITGRSGEIGDIYVEHEEVNFIN 227
 Db 194 LASGTVTVKPSSELTPLTALAALALQAGVPPGALNVVNGNAPEIGDALITSPQVRKIT 253

QY 228 FTGSPVQORIGKLAG--MRPIMLELGGKAGIIVLADADLONAAKQIVAGAYDSGQRCT 285
 Db 254 FTGSTAVGKKLMAAAAPTYYKKVUSLELGNAPSIVPDADLVAVKGTAAKTFNSGQICV 313
 QY 286 AIKRVLVVEEVADELAEKISENAVAKLSVGDPPF-DNATVTPVIDNSADPIESLVVDARQK 344
 Db 314 CANRVLVQDGIYKFAEAFSEAVQKLEVGDFRDTGTQGLINDAAVQKVFVQDAVSK 373
 QY 345 GAKELNEEKDGRLLT---PGLFDHVTLDKLAWEPEFGPILPIIRVKDAEBAVAIAKNS 401
 Db 374 GAKILIGKRHSGLMTFYPTVIRVDSDNMIMKEEIFGVPAPLIRFKTEEDAIRANDT 433
 QY 402 DFGLOSSVFTDFOKAFDIANKLEVTGTHINNKTRGPDNFFFLGKSGAGVQGIKXIS 461
 Db 434 IAGLAAYITFNSVQSRVFEALEVGLVGVNEGL-ISTEVAPFGVQKSGLSGREGSKYGM 492
 QY 462 EAMTVKSVIL-DM 474
 Db 493 DEYLEIKYVCLGDM 506
 RESULT 42
 A40872
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 5 precursor, mitochondrial - human
 C:Species: Homo sapiens (man)
 C:Date: 27-Mar-1992 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
 C:Accession: A40872
 R:Hsu, L.C.; Chang, W.C.
 J. Biol. Chem. 266, 12257-12265, 1991
 A:Title: Cloning and characterization of a new functional human aldehyde dehydrogenase ge
 A:Reference number: A40872; MUID:91286241; PMID:2061311
 A:Accession: A40872
 A:Molecule type: DNA; mRNA
 A:Residues: 1-517 <RSU>
 A:Cross-references: GB:M63967; NID:G337184; PIDN:AAA96830.1; PID:gi263008
 A:Note: 86-Ala and 107-Leu were also found
 C:Comment: Based on similarity, this form is assumed to be mitochondrial and to have pro
 C:Genetics:
 A:Gene: GDB:ALDH5; ALDHX
 A:Cross-references: GDB:I28788
 A:Map position: 9p13-9p13
 A:Note: the only intron occurs before the initiator codon
 C:Complex: homotetramer (probably)
 C:Function:
 A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreductase
 F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TNP>
 F:21-517/Product: aldehyde dehydrogenase (NAD+) 5 #status predicted <MAT>
 F:75-339/Domain: aldehyde dehydrogenase homology <ALDD>
 F:285,319/Active site: Glu, Cys #status predicted
 Query Match 27.6%; Score 661; DB 1; Length 517;
 Best local similarity 34.6%; Pred. No. 3e-35;
 Matches 169; Conservative 81; Mismatches 201; Indels 38; Gaps 11;
 QY 6 QNYVNGEWSVNVQ--TEILSPIDDSISLGFVPAMTVEEDVHAKGREAL---PAWALT 60
 Db 38 QLFINNEQDVAASKTFPTPTTGTGEVGHVAGDRAVDRAVKAAREVPRLSPPRRMD 97
 QY 61 VYERQYLKHAADIIRDEKEEIAIV---LAKEISKAYNASVTEVVRTADLIRYAAEERI 116
 Db 98 ASERGLNRLADLIVERDRVLASLETLDNKGPFQBSYALDDEVIK---VYRYFA--- 150
 QY 117 RLSTSADE-GKMADSTGHKLAVIRRPQVGIIVAIAPYVFNLSGSKIAPALIGNVVM 175
 Db 151 ---GWADKWHGKTIPMDQGHFFCTRHEPVGVCGLIIPWNPFLVMQGWKLAPALATNTV 207
 QY 176 FKPTGSGVSLVLAFAAGLPAAGVNTITGRSGEIGDIYVEHEEVNFINFTGTPVG 235

Db 208 MKVAEQPLSALYLASLIKEAGFPFGVWNIITGYGTAGAAIAQHMDVMDKVAFTGSTEVG 267
 QY 236 QRIKLAG--MRPTMLELGGKAGIVLADADLONAAKQIVAGADYDSQRCATKEVLV 292
 Db 268 HLIQKAGDSNLRVITLGGKSFIVLADADMEHAEVQCHEALFFNMGQCCAGSTFV 327
 QY 293 VEEVADLAELKISNVAKLSVGDFPDNATVT--PVIDNSADFTESLVVDARQKAKEL-- 349
 Db 328 EESIYNEFLERTVEKAKQKGVNPFELDTQQGFQVDKEQFERVLGYIQLGQKEGAKLGG 387
 QY 350 -NEFKDGRLLTPGLFDHVLDMKLAWEPFGPILPIRVKDAEBAVAIAKNSDFGLQSS 408
 Db 388 GERFGERGFIKPTVFGGVQDMRIAKEBIFQVPLFKFKTEEEVERANNTYGLAA 447
 QY 409 VTRFDKAFADIANKLVGTVHNNKTGRGPNF-----PFLGLKSGAGVQGIKYSIEA 463
 Db 448 VTRDLDDKAMYTQALQAGTWNVY-----NIVTCHTFPGFKESNGRELGEDGLKA 501
 QY 464 MTNVKSTVL 472
 Db 502 YTEVKVTVI 510
 RESULT 43
 E49343
 A:Title: aldehyde dehydrogenase (EC 1.2.1.-) - Pseudomonas sp. (strain C18)
 C:Species: Pseudomonas sp.
 C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 11-Jun-1999
 C:Accession: E49343; S27636
 R:Denome, S.A.; Stanley, D.C.; Olson, E.S.; Young, K.D.
 J. Bacteriol. 175, 6890-6901, 1993
 A:Title: Metabolism of dibenzothiothiophene and naphthalene in Pseudomonas strains: complete
 A:Reference number: A49343; MUID:94042852; PMID:8226631
 A:Accession: E49343
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-483 <DEN>
 A:Cross-references: GB:M60405; NID:g1511195; PIDN:AAA16129.1; PID:g294355
 C:Genetics:
 A:Gene: doxH
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase
 Query Match 27.6%; Score 660; DB 2; Length 483;
 Best Local Similarity 34.9%; Pred. No. 3.2e-35;
 Matches 166; Conservative 79; Mismatches 216; Indels 14; Gaps 6;
 QY 8 YVNGEKSSUNQ--IEILSPIDSSLSGFPVPMTRREEDVHAMKAGREALPAWALTIVERA 65
 Db 6 FINNAIDSSDQOTFERIHPVSSDVVTESANATVTDIAKAAQAAEAFKTKWAGFSERR 65
 QY 66 QYLKKAADIIRDKKEIATVLAKESKAYNASVTEVVRTADLRVAAEEGIRLSTSDRG 125
 Db 66 RLLKLVADWESKTPFKFEVNAVEGASALWAGNVASANVREAS----IATQI-QG 120
 QY 126 GKMDASTGHLKAVIRPQVIGVILAIAPYNYPNLSGSKTAPALIGCNVMPKPTQGSVS 185
 Db 121 ETIPTDKAETLSMTLRQVGPILSIVPNMGTAVALAARAIYPLVCGNTVNFKSGESRPA 180
 QY 186 GLVLAKAFAPAGLVNITG---RSGEIGDIVVHEEVNFINFTGTPVGQRIGKIA 242
 Db 181 HALITQCVQEGALPAGVLNINSPPDRGPEIADALISAKEIRINFTGSTRVGSIIAQA 240
 QY 243 G--MRPTMLELGGKAGIVLADADLONAAKQIVAGADYDSQRCATKEVLVVEEVADEL 300
 Db 241 AQHKLCLLELGGKPLVLDDADIDAIAKAAVFGSLFQGIQCMSTELLVDEKTADEF 300
 QY 301 AEKISENVAKLSVGDPP--FDNATVTPVIDNSADFTESLVVDARQKAKELNEFKDGR 358
 Db 301 VAKFEVKTRELKAGDPCTGDCIIGPMVSPNSGERINGLFKALDKKAVCGGLAQGAL 360
 QY 359 LTPGLFDHVTLMKLAEEFPGLPIPIRVKDAEBAVAIAKNSDFGLQSSVTRDFQKAF 418

Db 361 MPATILDHVKSDMRIYDEETFGPITVTVIRCKGAEAVRIANDSVYGLSSGVFGRDINEAL 420
 QY 419 DTANKLEVGTVHNNKTGRGPNFPLGLKSGAGVQGIKYSIEAMTNKSVILD 473
 Db 421 RVGMSIEYGVSHINGSTVQNEAQAQPYGGTKNTGYGRFGRVIDEFTFKWLTIE 475
 RESULT 44
 S15181
 A:Title: aldehyde dehydrogenase (EC 1.2.1.8) - Escherichia coli (strain K-12)
 N:Alternate names: betB protein
 C:Species: Escherichia coli
 C>Date: 09-Oct-1992 #sequence_revision 31-Jan-1997 #text_change 01-Mar-2002
 C:Accession: S15181; JQ1230; H64757; S10900
 R:Lamarck, T.; Kaasen, I.; Eshoo, M.W.; Falkenberg, P.; McDougall, J.; Strom, A.R.
 Mol. Microbiol. 5, 1049-1064, 1991
 A:Title: DNA sequence and analysis of the bet genes encoding the osmoregulatory choline-
 A:Reference number: S15178; MUID:92065800; PMID:1956285
 A:Accession: S15181
 A:Molecule type: DNA
 A:Residues: 1-490 <LAM>
 A:Cross-references: EMBL:X52905; NID:g48714; PIDN:CAA37092.1; PID:g48718
 A:Note: Part of this sequence, including the amino end of the mature protein, was confirm
 R:Boyd, L.A.; Adam, L.; Pelcher, L.E.; McHughen, A.; Hirji, R.; Selvaraj, G.
 Gene 103, 45-52, 1991
 A:Title: Characterization of an Escherichia coli gene encoding betaine aldehyde dehydroge
 A:Reference number: JQ1230; MUID:91348527; PMID:1875697
 A:Accession: JQ1230
 A:Molecule type: DNA
 A:Residues: 1-231, R', 233-311, P', 313-490 <BOY>
 A:Cross-references: GB:M77739; NID:g145403; PIDN:AAA23505.1; PID:g145404
 A:Experimental source: strain K-12
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: H64757
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-490 <BLAT>
 A:Cross-references: GB:AE000138; GB:U00096; NID:g1786501; PIDN:AAC73415.1; PID:g1786504;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: betB
 A:Map position: 7.5 min
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of betaine aldehyde to betaine using NAD+ and water
 A:Pathway: betaine biosynthesis
 A:Note: betaine is a protective osmolyte induced to accumulate by osmotic stress
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alternative initiators; homotetramer; NAD; oxidoreductase; stress-induced pro
 F:2-490/Product: betaine-aldehyde dehydrogenase #status experimental <NAT>
 F:45-306/Domain: aldehyde dehydrogenase homology <ALDD>
 F:252,286/Active site: Glu, Cys #status predicted
 Query Match 27.5%; Score 658.5; DB 1; Length 490;
 Best Local Similarity 31.9%; Pred. No. 4.1e-35;
 Matches 157; Conservative 101; Mismatches 199; Indels 35; Gaps 9;
 QY 4 EYQNVNVGKSSVN--QIETLSPIDSSLSGFPVPMTRREEDVHAMKAGREALPAWALTIV 61
 Db 6 EQLYIHGGVTSATSGRTFFETNPANGVLAIVQAGREDDVRAVSKAQCGKIWSMTA 65
 QY 62 YERQYLHKAADIIRDKKEIATVLAKESKAYNASVTEVVRTADLRVAAEEGIRLST 120
 Db 66 MERSRILRRAVDILREERDELAKLETLDTGKAYSETSVDIIVTGADVLEYVA----- 117
 QY 121 SADSGKMDASTGHLKAV-----ITROPVGVILAIAPYNYPNLSGSKTAPALIGNV 173
 Db 118 -----GLIPALEGSGIPLRETFSFYTRREFLGVAGIYQAWNYPIQIALWKSPALAAAGNA 172

A:Molecule type: protein
A:Residues: 2-501 <HEM>
R:Agarwal, D.P.; Cohn, P.; Goedde, H.W.; Hempel, J.
Enzyme 42, 47-52, 1989
A:Title: Aldehyde dehydrogenase from human erythrocytes: structural relationship to the :
A:Reference number: A61010; MUID:8937753; PMID:2776714
A:Accession: A61010
A:Molecule type: protein
A:Residues: 23-27;79-85;101-107;114-128;132-142;144-154;309-319;421-434;477-483 <AGA>
A:Experimental source: erythrocytes
R:Abrilla, D.P.; Fields, R.; Stein, S.; MacKerell Jr., A.D.; Pietruszko, R.
Biochemistry 26, 5679-5684, 1987
A:Title: Active site of human liver aldehyde dehydrogenase.
A:Reference number: A26817; MUID:88050861; PMID:3676276
A:Accession: A26817
A:Molecule type: protein
A:Residues: 266-273 <ABR>
C:Genetics:
A:Gene: GDB:ALDH1
A:Cross-references: GDB:l19667; OMIM:100640
A:Map position: 9q21-9q21
C:Complex: homotetramer
C:Function:
A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A>Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; cytosol; homotetramer; liver; NAD;
F:2-501/Product: aldehyde dehydrogenase (NAD+) 1 #status experimental <MAT>
F:59-323/Domains: aldehyde dehydrogenase homology <ALDD>
F:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F:269, 303/Active site: Glu, Cys #status experimental
F:456/Binding site: NAD (Cys) #status experimental

Query Match 27.5%; Score 658; DB 1; Length 501;
Best Local Similarity 34.1%; Pred.No. 4.5e-35;
Matches 163; Conservative 85; Mismatches 210; Indels 20; Gaps 8;

QY	8	YVNGEKSVN-QIEILSPIDSSIGFVPWATREEDVDHAKGAREAL---
Db	24	FINNEHDSVGKKAFPPNPATEEEELQCVEGSKEDVDKAVKAAQAFQISPTWTDAS 83
QY	63	ERAQLHKRAADIIERDKBEIATLAKESIKAY-NASVTEVVVTADLIKYAEEGIRLS 121
Db	84	ERGRLLYKLADIIRDRLMLLATMESMNGKLYSNAYLNLDLAGCIKTRYCAGWADKIQ-- 141
QY	122	ADEGGKMDASTGHKLAVIRROPQGVILAIPNYPNVLSCSKIAPALICGNVWFKPQTQ 191
Db	142	---GRTPIDGNFTTTRHEPIGVCGQIIIPWNPLVLMILWKIGPALSGCNVTVVKPAEQ 197
QY	182	GVSGLVLAKAFAGAPGAFVNFTITGRSGSIGDIYIVEHEEVNFINTGSTPVGQRIGKL 241
Db	198	TPLTALHAVSLIKEAGFPFNVPVGYGTAGAAISSHMWDKVAFTGSTVEGKLKEA 257
QY	242	AG---MRPIMLELGKGDAIVLADADLNAKQIVACADYSGQRCITAIRKVIIVVEVAD 298
Db	258	AGKSNIKRVTTELGGKSPCTVLADADLONAVEFAHGCVFYHQGCCIAASRFVEESIYD 317
QY	299	ELAEEKISENVAKLVSGDPFF-DNATVTPVIDNSADFTESLAWDAQKAK--ELNFKR 354
Db	318	EFRVRSVERAKYILGNPLTPGVTPQPQIDKEQYDKILDIESGKKGAOKLECQGPWGN 377
QY	355	DGRLLTPGLFDHVTLDMKLAEEPFGPILPIIRVKDAEBAVAIAKSDFGIQSSVFTRDF 414
Db	378	KGYFQVPTVFSNVTDENRIAKEEIFGPVOQIMKFSLDDVIKRVANTFYGLSAGVFTKOI 437
QY	415	QKAFDIANKLEVGVTHNNKTGRGPDNPFPLLGLKSGCAGVQGIRYSIEAMTNKYSIVL 472
Db	438	DKRAITTSALQAQGVVWV-NCYGVWSAQCPFGFRMSGNGRELGEYGFHEYTEKVTIVT 494

RESULT 46
95964

probable aldehyde dehydrogenase protein (EC 1.2.1.1-) [imported] - Sinorhizobium meliloti
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: C95964
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A>Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: C95964
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49379.1; PID:gl5140865; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMO21539
 A:Genome: plasmid
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 27.4%; Score 656; DB 2; Length 485;
 Best Local Similarity 36.1%; Pred. No. 5.8e-35;
 Matches 171; Conservative 84; Mismatches 201; Indels 18; Gaps 8;

QY 9 VNGEWKSSVNOEILSPIDSSSLGFVPAMTREVVDHAKGREALPAAALTVVERAQL 68
 DB 17 IAGLVDTDDREVRYPWNTVGVTPA---GRAHAREAPALAAAYOPKUTRYERQKIL 73
 QY 69 HKAADIIERKKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLSTSADEGGKM 128
 DB 74 LATAEALAAKKEISDVITLGLSKADSLVEVGRAFDVFTLAGQMCIR-----DGEIF 128
 QY 129 DAS-TGKLA---VIRQPGVIGVIAIPYVNLSSGKIAPALIGNVVMFPPTQGSV 184
 DB 129 SCDLTTPHGKARKITFMEPLTASAITFPNHLNVAHVAPAIATNNVNVVPTLTPM 188
 QY 185 SGLVLAKAPAGAGLPAGVNTITGRSGSIGDYIVHEVEVNFNTGSPVPGQIGKLAGM 244
 DB 189 TALLADILYENGUPPEMLSVTGWPDIGMEMTTPHVDLVTFTGSPVVKLIANAHY 248
 QY 245 RPIMLELGGKAGIV---LADADLNAKQIVAGADYSGQRTAIRKVLVVEEVADELA 301
 DB 249 KRQVLEGGNDPLIILNDLSDDDLARAADLAVAGATKNSQRCRTAVKRIILCQSVADRFV 308
 QY 302 EKISENVAKLSVGGPFDNAT-VTPVIDNSADFTESLVVDARQKAKELNFRDRLIT 360
 DB 309 PLVIERAKLRFQDPMDRSITDGTVTHEKAAALFEERVMAEEGADILYHPCRSALLP 368
 QY 361 PGLFDRVTLDMKLAWEPPGPIPIIRVKDAEEA-VAIAKNSDFGLQSSVFTDFOKAFD 419
 DB 369 PIVDRVPHQSDVLVEFTGPIPIVVRPDDDDATITLNSFARGLSSGVCTNDYRMQK 428
 QY 420 IANKLEVGVTHNNKTGRGPNFPFLGLKSGAGV-QGIRYSIEAMTNVKSIVL 472
 DB 429 YIAGLVKGVTVNTEVPGVRIENSPFGIKDGSNGYKGVIEAKMFTNVTATSL 482

RESULT 47

AD3300

succinate-semialdehyde dehydrogenase [NAD(P)] (EC 1.2.1.16) [imported] - Brucella melite
 C:Species: Brucella melitensis
 C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-May-2002
 C:Accession: AD3300
 R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,

; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessc
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis;
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AD3300
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-487 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AL51567.1; PID:gl7982287; GSPDB:GN00190
 A:Experimental source: strain 16W
 C:Genetics:
 A:Gene: BMEI0386
 A:Map position: 1
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 27.4%; Score 655; DB 2; Length 487;
 Best Local Similarity 35.3%; Pred. No. 6.8e-35;
 Matches 169; Conservative 81; Mismatches 201; Indels 28; Gaps 11;

QY 6 QNYVNGEWKSSVN--QIILSPIDSSSLGFVPAMTREVVDHAKGREALPAAALTVYE 63
 DB 13 QCLVNGRWIDADGTTIKVTNPADGSGVIGTVPVSLSVATIKEDASAKALSQWAAKTAKE 72
 QY 64 RAQYLHKAADIIERKKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLSTSD 123
 DB 73 RAGILRKWFDLIIANADDIALIMTSGQKPLAEARGEVLVYAAFIENFAEAKRV----- 127
 QY 124 EGGWMDA-STGHKLAVIRQPGVIGVIAIPYVNLSSGKIAPALIGNVVMFPKPTQG 182
 DB 128 YGDTIPAPQNGORLTVI-ROPVGVTAATTPNFPAAITTKAAPALAAAGCTMIVRPADLT 186
 QY 183 SVSGLVLAKAFAGLPAGVNTITGRSGSIGDYIVHEVEVNFNTGSPVPGQIGKLA 242
 DB 187 PTLALAILLAKAGIPAGVLIQVTKAREIGAEITSNDTVKLSFTGSTEVRLL--MA 244
 QY 243 GMRP----IMLELGGKAGIVLADADLNAKQIVAGADYSGQRTAIRKVLVVEEVD 298
 DB 245 QCARTIKRISLELGGNAPFIVFDADLDAADVGMVSKYRNAGQTCVCANRIYVQGVYD 304
 QY 299 ELAKISENVAKLSVGGPFDNATVTPVIDNSADFTESLVVDARQKAKELNFRKDRGR 357
 DB 305 KFAEKLAARKVGLKVGNGTEPGVVGIVGEMIEEKATIKVAHIEDAVSKGAKLITGGKELGG 364
 QY 358 L-LPFGLEFHDVTLDMKLAWEPPGPIPIIRVKDAEEAVALANKSDPGLQSSVFTDFQX 416
 DB 365 LPFEGILITGVTSMLVAKETFGPIAPLAFADTIEEVEVIAQANDTIFGLAAVFTIENFSR 424
 QY 417 AFDIANKLEVGVTV-H---INNKTRGPDNFPFLGLKSGAGVQGGIRYSIEAMTNVKS 470
 DB 425 AIRVSEALEYGVMGVHNTGLISNEVA-----PFGGVKQSLGREGSKYIEEYLETXYI 477

RESULT 48

B85524

NAD+-dependent betaine aldehyde dehydrogenase [imported] - Escherichia coli (strain O157:
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B85524
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B85524
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <STO>
 A:Cross-references: GB:AE005174; NID:gl2513124; PIDN:AA054654.1; GSPDB:GN00145; UWGP:Z039
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: betB
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 27.3%; Score 653.5; DB 2; Length 490;
 Best Local Similarity 32.4%; Pred. No. 8.6e-35;
 Matches 157; Conservative 100; Mismatches 207; Indels 21; Gaps 9;

QY 4 EYQYVNGEWSKSVN--QIEILSPIDSSISLGFVPAMTREVVDHAKAGREALPAAALTV 61
 DB 6 EQQLYHGGYTSATGRFTETINPANGVLATVQAGREVDRAVSKAQOQKIIWAAMTA 65
 QY 62 YERQYLKKAADIIRDEKEELATVLAKEISKAYN-ASVTEVVRTADLIRYAAEGIRLST 120
 DB 66 MERSRILARRAVDILRERDELAKLETIDTGKAYSTSTVIVTGADVLEYA--GLIPSL 123
 QY 121 SADEGGKMDASTGHKLAIVIRQPVGIVLAIPYNPVNLSSKIAPALIGGNVVMFKPPT 180
 DB 124 EGSQIPURETS----FVTRREPLGVVAGIGAWNPYIQIALWKSAPALAGNAMIKPS 179
 QY 181 QGSVSGVLAKAPAFAGIACAGVNTITGRSGISGIDYIVHEEVNFNFTGTPVQRI-- 238
 DB 180 VTPIULKLAETYSAGLPDGVNLPVGVGAETGQYLEHFGIAKVSFTGGVASKKWA 239
 QY 239 -GKLAMRPIMLELGGKDGAGIVLADADLONAAKQIVAGDYSGQRTAKRVLVBEVA 297
 DB 240 NSASSLKEYTMELGKSPFLVDDADLDLADIAWMANFSSGVQCTNGTRVFPACKN 299
 QY 298 DELAEKISENVAKLSVGPFDNAT-VTPVIDNSADFISSLVVDARQKAKEL---NEFK 353
 DB 300 AAFEQKILARVERIRAGVDPQTNFGPLVSPFRDNLVRYIAKKEGARVLGGDVLK 359
 QY 354 RD---GRLTLTGLFDHVTLDKLAWEPPGPIPLIRVKDAEVAIVANKSDGLQSSVTR 409
 DB 360 GDGLDNGAWVAPVTFDSCDEMTIVRETFPGVNSILTYSEDEVIIRANDTDYGLAAGI 419
 QY 410 FTRDFQKAFDIANKLEVTGHNNKTKRGDPNFPFLGLKSGAGVQGIYRISIEAMTNVKS 469
 DB 420 VTADLNEARHVIHQLEAGICWINT-WCESPAEWPGVGVKSGIGRENGWMLQSVTVQKS 478
 QY 470 IVDLM 474
 DB 479 IQVEM 483

RESULT 49

A95990
 Probable dehydrogenase, possibly aldehyde dehydrogenase protein [EC 1.2.1.-] [imported]
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: A95990
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 96, 9889-9894, 2001
 A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95942; MUID:21396508; PMID:11481431
 A:Accession: A95990
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-477 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CA49585.1; PID:gl5141072; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 233, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 Hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMD20891
 A:Genome: plasmid
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 27.3%; Score 653; DB 2; Length 477;
 Best Local Similarity 33.8%; Pred. No. 8.9e-35;
 Matches 161; Conservative 78; Mismatches 219; Indels 18; Gaps 7;

QY 5 YQYVNGEWSKSVNQIEILSPIDSSISLGFVPAMTREVVDHAKAGREALPAAALTVYER 64
 DB 4 HQNLTAGVGGGVANINPNTDVVGEYARASAEADAKAIAAFAAPFAWRSRSGILER 63
 QY 65 AQYLKKAADIIRDEKEELATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLSTSADE 124
 DB 64 HAILKKTADLARKDELGRLLSRECKTIAEGIGETVRACQIEFFPAGETURLA----- 118
 QY 125 GKMDDASTGHKLAIV-IRQPVGIVLAIPYNPVNLSSKIAPALIGGNVVMFKPPTGGS 183
 DB 119 -GEVVPVSRPGIVGIVETREPAGVGLITPWNFFIAPAKLAPALCYGNTIVFKDAELVP 177
 QY 184 VSGVLAKAFABAGLPAGVFNITITGRSGEIGDYIVHEEVNFNFTGTPVQOR--IGKL 241
 DB 178 GCSWAIVDILHRAGLPKGVLLNWGKSVVGGQMLDSDPDVQAITFTGSTATGRVAVASV 237
 QY 242 AGMRPIMLELGGKDGAGIVLADADLONAAKQIVAGDYSGQRTAKRVLVVEVADELA 301
 DB 238 EHNRYQLEMGKKNPFVLLDADLSVAEAVNSAFFSTGQRTASSRIIVTEGIDHDFV 297
 QY 302 EKISENVAKLSVGPFDNAT-VTPVIDNSADFISSLVVDARQKAK-----ELNEFKRD 355
 DB 298 AAMGERIKGLVVDALKPQTHIGFVVDQSQLNODTYIAIGKQEGAKLAFAGGEVISRDP 357
 QY 356 GRLTTPGLFDHVTLDKLAWEPPGPIPLIRVKDAEVAIVANKSDGLQSSVTRDFQ 415
 DB 358 GYILOPALFTENENKISRSEIFGPVAIVIRVKDYDEALAVANDPFLUSSGIATSLK 417
 QY 416 KAFDIANKLEVTGHNNKTKRGDPNFPFLGLKSGAG--VQGIYRISIEAMTNVKS 469
 DB 418 HATHEKRNAGWVMVNLPTAGVDFHVFPGKASGYPREQG-KYAAEFYNVKT 472

RESULT 50

H72562
 Probable aldehyde dehydrogenase APE1786 - Aeropyrum pernix (strain KI)
 C:Species: Aeropyrum pernix
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: H72562
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ki
 DNA Res. 6, 83-101, 1999
 A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyri
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: H72562
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <KAW>
 A:Cross-references: DDBJ:AP000062; NID:gs105244; PIDN:BA80789.1; PID:gs105476
 A:Experimental source: strain KI
 C:Genetics:
 A:Gene: APE1786
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 27.2%; Score 651.5; DB 2; Length 485;
 Best Local Similarity 33.9%; Pred. No. 1.1e-34;
 Matches 163; Conservative 93; Mismatches 210; Indels 15; Gaps 7;

QY 5 YQYVNGEWSKSVNQIEILSPIDSSISLGFV----PAMTREVVDHAKAGREALPAAALT 60
 DB 2 FKTIVAGNVLTSLNLAIVRSPIDGSIIAKVSRDAAALEEGVDVWVRLGRRRI---RDVP 58
 QY 61 VYERQYLKKAADIIRDEKEELATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLST 120
 DB 59 GEKRLNLSRIADLLEKNGAGDFEVLTVNAGTRKQAAGEVEASIDRLRKAALD-LRLK 117
 QY 121 SADEGGKMDASTGHKLAIVIRQPVGIVLAIPYNPVNLSSKIAPALIGGNVVMFKPPT 180
 DB 118 GRYVPGWDRHTLESEGIVRREPVCVVLIIFNFTPLDFTVNFVSTIPGNAFTVKEPS 177

Search completed: April 13, 2004, 14:22:58
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:22:34 ; Search time 47 Seconds
(without alignments)
2657.352 Million cell updates/sec

Title: US-09-868-195-12

Perfect score: 2393

Sequence: 1 LFKEYQNVNGEWKSVNQI.....GIRYSIEANTVKSIVLDMK 475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues

Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1802	75.3	474	12	US-10-282-122A-73974
2	1700	71.0	475	12	US-10-282-122A-72316
3	1693	71.0	439	9	US-09-815-242-13316
4	1639	68.5	475	12	US-10-282-122A-74625
5	1397	58.4	486	12	US-10-282-122A-53543
6	1388.5	58.0	486	12	US-10-282-122A-52739
7	1383	57.8	482	15	US-10-282-122A-52073
8	1335	55.8	472	15	US-10-369-493-16536
9	1312	54.8	473	15	US-10-369-493-16664
10	1244	52.0	481	15	US-10-369-493-17301
11	1038.5	43.4	503	12	US-10-424-599-187423
12	1001.5	41.9	434	12	US-10-424-599-187419
13	868.5	36.3	475	12	US-10-282-122A-76821
14	786.5	32.9	492	14	US-10-282-518-4
15	777	32.5	493	9	US-09-823-901-9

16	777	32.5	493	14	US-10-175-696-21
17	775.5	32.4	492	14	US-10-172-585-13
18	763	31.9	488	15	US-10-369-493-23067
19	759.5	31.7	463	15	US-10-369-493-1073
20	741.5	31.0	455	15	US-10-369-493-1188
21	741.5	31.0	484	9	US-09-738-626-6430
22	737	30.8	476	15	US-10-369-493-17
23	735	30.7	482	15	US-10-369-493-17249
24	728	30.4	470	15	US-10-369-493-16454
25	726	30.3	482	9	US-09-815-242-10264
26	726	30.3	482	12	US-10-282-122A-56650
27	726	30.3	482	15	US-10-369-493-23511
28	721.5	30.2	483	12	US-10-282-122A-46932
29	718	30.0	488	12	US-10-282-122A-60838
30	716.5	29.9	493	15	US-10-369-493-18199
31	713.5	29.8	480	15	US-10-369-493-18593
32	711	29.7	482	12	US-10-282-122A-59544
33	710.5	29.7	489	15	US-10-369-493-18711
34	705.5	29.5	473	15	US-10-369-493-13571
35	705.5	29.5	491	9	US-09-997-664-3
36	704	29.4	480	15	US-10-464-952-3
37	697.5	29.1	468	12	US-10-282-122A-68209
38	694	29.0	473	15	US-10-369-493-17169
39	694	29.0	477	15	US-10-369-493-18996
40	694	29.0	477	15	US-10-369-493-8741
41	693	29.0	480	12	US-10-282-122A-69547
42	693	29.0	480	12	US-10-282-122A-70076
43	692	28.9	482	12	US-10-282-122A-75217
44	691	28.9	478	15	US-10-369-493-13957
45	690	28.8	489	12	US-10-282-122A-51224
46	687	28.7	480	15	US-10-369-493-11184
47	686	28.7	467	15	US-10-369-493-8312
48	682.5	28.5	458	15	US-10-369-493-4282
49	682.5	28.5	458	15	US-10-369-493-7037
50	682.5	28.5	485	15	US-10-369-493-17279

ALIGNMENTS

RESULT 1

US-10-282-122A-73974
; Sequence 73974, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```
; PRIOR APPLICATION NUMBER: 60/253,625
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 73974
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73974

Query Match      75.3%; Score 1802; DB 12; Length 474;
Best Local Similarity 73.7%; Pred. No. 4.1e-151;
Matches 347; Conservative 58; Mismatches 66; Indels 0; Gaps 0;

QY 5 YQNVNGKSSVNOIETLSPIDDSSIGFVPAMTREVVDHAKAGREALPAWAALTYYER 64
Db 4 YQNVNGKSSVNOIETLSPIDDSSIGFVPAMTREVVDHAKAGREALPAWAALSATIER 63
QY 65 AQYLHKAADIIRDERKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSTSADE 124
Db 64 AAYLHKAADIIRDERKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSTQAME 123
QY 125 GKMDASTGHKAVIRPQVIGVLAIPYVPVNLGSKAPALIGGNVVMFKPTQGSV 184
Db 124 GGGFEAAKSKLVAVRREPVGVLAIAPENYVNLGSKAPALIGGNVVMFKPTQGS 183
QY 185 SGLVAKAFPAEAGLPAGVNTTITGRGSEIGDYIVVEHVEVNFINTGSPVQGRCKLAGM 244
Db 184 SGLLAKAFPAEAGLPAGVNTTITGRGSEIGDYIVVEHVEVNFINTGSPVQGRCKLAGM 243
QY 245 RPIMLELGGKDAAGIADADLDNAKQIVAGAYDSGQRTATKRVLVVEEVADELAEKI 304
Db 244 RPIMLELGGKDAAGIADADLDNAKQIVAGAYDSGQRTATKRVLVVEEVADELAEL 303
QY 305 SENTAKLSVGDPPDNATVTPVIDNSADFIESLVVDARQKAGKELNEFKRDLITPGLF 364
Db 304 QEVSKULTVGDPPDNATVTPVIDNSADFIESLVVDARQKAGKELNEFKRDLITPGLF 363
QY 365 DHVTLDMKLAWBEPGPILPIIRVKDAEAAVAIAKNSDFGLQSSVFTDFOKATDIANKL 424
Db 364 DQVTKDMKVABEPGPILPIIRVKDAEAAVAIAKNSDFGLQSSVFTDFOKATDIANKL 423
QY 425 EVGVTHNNKTGRGPDNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 475
Db 424 EVGVTHNNKTGRGPDNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 474
```

RESULT 2

```
US-10-282-122A-72316
; Sequence 72316, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUIPA.034A
```

```
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 72316
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72316
```

```
Query Match      71.0%; Score 1700; DB 12; Length 475;
Best Local Similarity 66.9%; Pred. No. 4.7e-142;
Matches 318; Conservative 75; Mismatches 82; Indels 0; Gaps 0;

QY 1 LTKYQNVNGKSSVNOIETLSPIDDSSIGFVPAMTREVVDHAKAGREALPAWAALT 60
Db 1 MTKQYQNVNGKSSVNOIETLSPIDDSSIGFVPAMTREVVDHAKAGREALPAWAALS 60
QY 61 VYERAAQYLHKAADIIRDERKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120
Db 61 YIERAAQYLHKAADIIRDERKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120
QY 121 SADEGKMDASTGHKAVIRPQVIGVLAIPYVPVNLGSKAPALIGGNVVMFKPT 180
Db 121 EVLEGGSFEAAKSKLVAVRREPVGVLAIAPENYVNLGSKAPALIGGNVVMFKPT 180
QY 181 QGSVSLGLVAKAFPAEAGLPAGVNTTITGRGSEIGDYIVVEHVEVNFINTGSPVQGRCK 240
Db 181 QGSISGLLAKAFPAEAGLPAGVNTTITGRGSEIGDYIVVEHVEVNFINTGSPVQGRCK 240
QY 241 LAGMRPIMLELGGKDAAGIADADLDNAKQIVAGAYDSGQRTATKRVLVVEEVADEL 300
Db 241 MAGMRPIMLELGGKDAAGIADADLDNAKQIVAGAYDSGQRTATKRVLVVEEVADEL 300
QY 301 AEKISENVAKLSVGDPPDNATVTPVIDNSADFIESLVVDARQKAGKELNEFKRDLIT 360
Db 301 VEKIREKVLATITGNPEDDADITPLIDTSADYVEGLINDANDKGAALTEIKREGALIC 360
QY 361 PGLFDHVTLDMLKAWBEPGPILPIIRVKDAEAAVAIAKNSDFGLQSSVFTDFOKAFDI 420
Db 361 PLEFDKVTITDMRLAWBEPGPILPIIRVKDAEAAVAIAKNSDFGLQSSVFTDFOKAFDI 420
QY 421 ANKLEVGTVHNNKTGRGPDNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 475
Db 421 AEQLEVGTVHNNKTGRGPDNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 475
```

RESULT 3

```
US-09-815-242-13316
; Sequence 13316, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
```

; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13316
 ; LENGTH: 439
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-815-242-13316

Query Match 71.0%; Score 1699; DB 9; Length 439;
 Best Local Similarity 74.5%; Pred. No. 5e-142;
 Matches 327; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

 QY 37 MTREVDHAKKAGREALPAWALTVERAQLHKAADIIERDKEEATVLAKEISKAYNA 96
 Db 1 MTQTEADFAQARAALPAWALSALERAALYHKAALIERDKEEATVLAKEISKAYNA 60

 QY 97 SVTEVTRTADLIRYAAEEGIRLSSTADEGKMDASTGHKLAVIRROQVGVILAIAPNYP 156
 Db 61 AIGEVTRTADLIRYAAEEGIRITGOMEGGFEASKNKLAVRREPVGILAIAPNYP 120

 QY 157 VNLSSKIAPALIGNVNMFPTQGSVGLVIAKAFABAGLPAGVNTITGRSEIGDY 216
 Db 121 VNLSSKIAPALIGNVNMFPTQGSISGLLLAKAFEEAGIPAGVNTITGRSEIGDY 180

 QY 217 IVEHEEVNFTGTPVQGRIGKLAGMRPIMLEGGKDGIVLADADLNAKQIVAGA 276
 Db 181 IIEHEVNFNFTGTPIGERIGLAGMRPIMLEGGKDALVLEDDLEHAKQIVAGA 240

 QY 277 YDYSQRCATKRVIVVEVADELAEKISENVAKLSVGPDPDNATVPVIDNSADFTES 336
 Db 241 FYSYQRCATKRVIVLESVADKATLLOEVSUKLVGDPDNATVPVIDNASADFTWG 300

 QY 337 LVVDARQKAGELNEFRDGRLLITFGLFDHVTLDMLKAWBPFPPIPIRVKDAEENVA 396
 Db 301 LIEDAQEKAQALTPIKREGNLLWFLFDQVTKDMKVAWBPFPPIPIRVASVEEAIA 360

 QY 397 IANKSDFGLQSVFTDFQAFDANKLEVGTVHNNKTGRGPDNRPFLGLGSGAGVQG 456
 Db 361 FANESFGLQSVFTDFQAFDANKLEVGTVHNNKTGRGPDNRPFLGLGSGAGVQG 420

 QY 457 IRYSTEAMTNKSVILDMK 475
 Db 421 IKYSIETAMTNKSVIFDVK 439

RESULT 4

US-10-282-122A-74625
 ; Sequence 74625, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 74625
 ; LENGTH: 475
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pyogenes
 US-10-282-122A-74625

Query Match 68.5%; Score 1639; DB 12; Length 475;
 Best Local Similarity 65.5%; Pred. No. 1.2e-136;
 Matches 311; Conservative 77; Mismatches 87; Indels 0; Gaps 0;

 QY 1 LTKYQNVYNGEWSKSNQIEILSPIDSSLGFPDAMTREVVDHAKAGREALPAWALT 60
 Db 1 MAKQXKNLVNGEWSKSNQIEILSPIDSSLGFPDAMTREVVDHAKAGREALPAWALT 60

 QY 61 YVERAQLHKAADIIERDKEEATVLAKEISKAYNASVTEVTRTADLIRYAAEEGIRLS 120
 Db 61 YVERAQLHKAADIIERDKEEATVLAKEISKAYNASVTEVTRTADLIRYAAEEGIRLS 120

 QY 121 SADEGKMDASTGHKLAVIRROQVGVILAIAPNYPVNLSSKIAPALIGNVNMFPT 180
 Db 121 EVLEGGSEASKKKTAIVRRPVGILVLAISPNYPVNLSSKIAPALIGNVNMFPT 180

 QY 181 QGSVGLVIAKAFABAGLPAGVNTITGRSEIGDYIVEHEEVNFTGTPVQGRIGK 240
 Db 181 QGSISGLLLAKAFEEAGIPAGVNTITGRSEIGDYIVEHEEVNFTGTPVQGRIGK 240

 QY 241 LAGMRPIMLEGGKDGIVLADADLNAKQIVAGADYDYGORCTAKRVIVVEVADEL 300
 Db 241 LAGMRPIMLEGGKDGIVLADADLNAKQIVAGADYDYGORCTAKRVIVVEVADEL 300

67	EMPVYKRANIFHKAADLLENIDISIANILVLVEIAKDIKSARAEEVETADFLRYTADYGN	126	
QY	118	LSTSADEGKMWDASTGHKLAVIRQPGVIGVLAIAPVNPVNLGSKIAPALIGTNVMEK	177
Db	127	MEGEAISGDNFPGGTNRKXSVSRVPLGTVLAIAPNPVNPWLSKIAPALIGNAVILK	186
QY	178	PPTQSGVSGVLVLAKAPAFAGLPGAVFNNTITGRSGEIGDYIHEEVEVNFINTGTPVGQR	237
Db	187	PATQCAISALHVVETMRKAGIPDGLVNTITGRSGEIGDYVVTTHKGINFINTGSTVEGQH	246
QY	238	ICKLAGVRPIMLELGCKQDAGIVLADADLDNAAKQIVAGADYSGQRCTAIKXVLVVEEVA	297
Db	247	ISKISGMVPLLEELGKQAAIIVLEDADLDFAAKNTVSGAYSYSQGRCTAVKRIIIVQESVA	306
QY	298	DELABKISENVAKLSVGDPFONATVPVTDNNSADFTESLAVDARQKAGKELNEFRDGR	357
Db	307	DKLVCKKPIVEKLTIGMPDEVVITPLIDNKATDFVOGLVDDALHKGAKUITGNVRKNN	366
QY	358	LTPGLEFHDVTLDMKIAEPEFGPIPIIRVKDAEAAIANKSDPGLQSSVFTDFEQA	417
Db	367	LFVPTLLDNVNDMKIAEPEFGPGVPIIRVKDINQALETANQSEVGLQSSVFTSDIDKA	426
QY	418	FOIANKLEVTGTHNNKTRGPDNFPPLGLKSGGAGVQGTIRYSIEMTNVKSIVLDMK	475
Db	427	FYIADKLEVTGTHNNKTRGPDHFFPLGVKASGNGTGVKYSIEMTRKAVVNVYR	484

97

```

QY      298 DELAEKISENVAKLISVGDFDNATVTPVIDDINSADFIETSLIYVDARQGAKELNEFKEDGR 357
DB      307 DKLVGKIPLVEKITGNPMDEVWITPLIDNKATDFVQGLVDDALHKGAKLITGNVRKNN 366
QY      358 LTTCLGDHYTLDMLKLAEEFPGILPIIRVKDAEEAVATANESDGLQSSVFTTFDQKA 417
DB      367 LFYPFLDLNVNDMLKLAEEFGGVLPPIRVKDINQALETANOSSEYGVQSSVFTSDIDKA 426
QY      418 FDIANKLEVGTVINHNKTGRGFDPNFPLGLKSGAGVGQIRYSIEAMTNVKSIVLDMK 475
DB      427 FYIADKLEVTGVIQNNKTERGEDHFPPFLGVKASGMGTGGVKYSIEAMTRPKAVVVNR 484

RESULT 6
US-10-282-122A-52739
; Sequence 52739, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITEA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 52739
LENGTH: 486

```

```

/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 52739
/ LENGTH: 486

```

; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52739

Query Match 58.0%; Score 1388.5; DB 12; Length 486;
Best Local Similarity 56.7%; Pred. NO. 2.2e-114; Indels 2;
Matches 268; Conservative 90; Mismatches 112; Gaps 3;

QY 5 YQYVNGEWKSSV--NQIEILSPIDSSLGTFVPAMTRVEVDHAMKAGREALPAWAAITVY 62
DB 12 YKFFGGQWNSKTNFIISPIDSSLGTFVPAMTRVEVDHAMKAGREALPAWAAITVY 71
QY 63 ERAQYIHKAAIIRDEKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLSTSA 122
DB 72 ERATILLKAANILDEKAEBIADIMTKETIAKDKKSSISEVVRTADYIRFSADTAKNVGET 131
QY 123 DGGKMDASTGHKLAIVRQPGVILALAPYVNVLSGSKIAPALIGNVVMEKPPPTQG 182
DB 132 LFGDFFGNSKGISIVNRPVGVVLSFPNPNLSGSKIAPALMAGNSVVLKAPTQG 191
QY 183 SVSGIVLAKAPAEAGLPAGVNTITGRGSEIGDYIVVEHEEVNFTGTPVQGRIGKLA 242
DB 192 SISALHLVYKVEBAGLPAGVNTITGRGSEIGDYIVSHPMIDFNTFGSTGVKRIHSIS 251
QY 243 GMRPIMLEGGKAGIVLADADLNAKQIVAGDYSGQRCTAIKRVLVVEVVADELA 302
DB 252 VMKPMLEGGKAAIVLEADLDLAAKNTVSGAYSGQRCTAVKRIILVBEIADTLVE 311
QY 303 KISENVAKLSVGDPF--DNATVTPVIDNSADFIESIIVDARQKAKELNEFKDGLLTP 361
DB 312 KYKRVESLKTGNPLEKEQVDIVPLIDDKAADYVEALMEANNGATLLVGGKREGNLIYP 371
QY 362 GLFDHVTLDMLKAWPEPFGPIPIIRVKDAEBAVAIAANKSDFGLOQSSVFTROPKAFDIA 421
DB 372 TLFDNVTLDMELAWPEPFGPIPIIRVKDQVDEAIRANESYEGLOASIFPTKEINAFYA 431
QY 422 NKLEVGTVHNNKTRGPDNPFPLGLKSGAGVQGIIRYSIEAMTVNKSIVLDM 474
DB 432 EKLEVGTVQVNNKPERGDPHPFGIGIKASGLGTQGIKYSIEAMSRPKAIVLNI 484

RESULT 7

US-10-282-122A-52073
; Sequence 52073, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52073
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-52073

Query Match 57.8%; Score 1383; DB 12; Length 482;
Best Local Similarity 57.0%; Pred. NO. 6.7e-114;
Matches 269; Conservative 86; Mismatches 115; Indels 2; Gaps 2;

QY 5 YQYVNGEWKSSVNO--IETLSPIDSSLGTFVPAMTRVEVDHAMKAGREALPAWAAITVY 62
DB 11 YKLFDFGKWSKTNKTIETHSPDGSLIGKVQALSKEVDIFKSSSTQCKWGETPIN 70
QY 63 ERAQYIHKAAIIRDEKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLSTSA 122
DB 71 ERATIRKAAIILODNAEYIAKILSNETAKDLKSSLSSEVKRTADTIFRTANEGTHMEGA 130
QY 123 DGGKMDASTGHKLAIVRQPGVILALAPYVNVLSGSKIAPALIGNVVMEKPPPTQG 182
DB 131 INSDNPPGSKDKLSEVERVPLGIVLALSPNPNVLSGSKVAPALAGNSVVLKPSITG 190
QY 183 SVSGIVLAKAPAEAGLPAGVNTITGRGSEIGDYIVVEHEEVNFTGTPVQGRIGKLA 242
DB 191 AISALHAEIIFNAAGLPAGVNTITGRGSEIGDYILTHEEVNFTGTPVQGRIGKLA 250
QY 243 GMRPIMLEGGKAGIVLADADLNAKQIVAGDYSGQRCTAIKRVLVVEVVADELA 302
DB 251 GMPVYVLEGGKAAIVLEADANLETTAKSIVSGAYSGQRCTAVKRVLMVDKVADELVE 310
QY 303 KISENVAKLSVGDPF--DNATVTPVIDNSADFIESIIVDARQKAKELNEFKDGLLTPG 362
DB 311 LVTKKELKVGPNFDDVTITPLIDNKAADYVQTLIDDAIEKGATLIIVGNKRKENIMYPT 370
QY 363 LFDHVTLDMLKAWPEPFGPIPIIRVKDAEBAVAIAANKSDFGLOQSSVFTROPKAFDIA 422
DB 371 LFDNVTLDMELAWPEPFGPIPIIRVKSMDEALANRSEYGLQSAVFTENMHDAYFIAN 430
QY 423 KLEVGTVHNNKTRGPDNPFPLGLKSGAGVQGIIRYSIEAMTVNKSIVLDM 474
DB 431 KLDVGTVQVNNKPERGDPHPFLGTGSSGMGTQGIIRYSIEAMTRHKSIVLNL 482

RESULT 8

US-10-369-493-16536
; Sequence 16536, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 16536
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-10-369-493-16536

Query Match 55.8%; Score 1335; DB 15; Length 472;
Best Local Similarity 55.7%; Pred. No. 1.2e-109;
Matches 263; Conservative 80; Mismatches 127; Indels 2; Gaps 1;

QY 5 YQNVNGWK--SSVNOIEILSPIDSSILGFVPAMTREVVDHAKAGREALPAWALTYY 62
DB 1 YKFLVNGWRESSSETIEIPSLHEVIGQVQAITCEVDEIAIAKAEQKSWAKASLQ 60
QY 63 ERAQYLHKAADIERDKEEATVLAKESIKAYNASVTEVVRTADLIRYAAEGIRLSTSA 122
DB 61 DRKLYLKWADELVNQDEIADIIMKEVGKGYKDAKEVVRTADFRITYIEEALHMHGES 120
QY 123 DEGGKMDASTGHKLAVIRROPVGIIVLAIPNYPNLSGSKIAPALIGNVVMFKPTQOG 182
DB 121 MWGDSFPGGTSKLAIIORAPLGVLAIAFPNYPNLSAAKLAPALINGNAVIFKPAATQ 180
QY 183 SVSGVLAKAFEAAGLPGAVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGORIGKLA 242
DB 181 AISGIRKVAUHLKAGLPGVNVATGRGSAIGDYLVEHGINNVSTFGTNTGKLAKTA 240
QY 243 GMRPIMLELGGKDGAGIVLADADLNAAKQIVAGAYDYSQORCTAIKRVLVVEEVADELA 302
DB 241 SMPLVLELGGKDPGIVREDADLQDAANHIVSGAFYSQORCTAIKRVLVHENVADELVD 300
QY 303 KISNVAKLSVGDPPDNATVTPVIDNSADFIESLVDARQKGAKELEFEDGRLITPG 362
DB 301 LLKAQVAELSVGSPQDSTIVRLDDKSADFFVQGLVDVAKEGATIVIGNKRENLITYPT 360
QY 363 LFDHVTLDMLKAWEPFGPIPIIRVKDAEEAVATANKSDFGLOQSVFTRDFOKAFDIAN 422
DB 361 LIDHVTMVKAWEPFGPIPIIRVSSDEQALIANKSEFGLOQSVFTKDKAKAFAN 420
QY 423 KLEVGTVHNNKTRGPDPNFFPLGLKSGAGVQGIIRYSIBAMTNVKSIVLDM 474
DB 421 KIETGSVQINGRTERGPDHPFFIGVKGSGWGAQGIKRSLSMTREKVTVLNFK 472

RESULT 9

US-10-369-493-16664
; Sequence 16664, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16664
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-10-369-493-16664

Query Match 54.8%; Score 1312; DB 15; Length 473;
Best Local Similarity 55.2%; Pred. No. 1.3e-107;
Matches 261; Conservative 77; Mismatches 133; Indels 2; Gaps 1;

QY 5 YQNVNGWK--SSVNOIEILSPIDSSILGFVPAMTREVVDHAKAGREALPAWALTYY 62
DB 1 YKFLVNGWRESSSETIEIPSLHEVIGQVQAITCEVDEIAIAKAEQKSWAKASLQ 60

DB 1 YNVYNGWRESSSETIEISSPYLHEVIGVQVQAITCEVDEIAIAKAEQKSWAKASLQ 60
QY 63 ERAQYLHKAADIERDKEEATVLAKESIKAYNASVTEVVRTADLIRYAAEGIRLSTSA 122
DB 61 DRKLYLKWADELVNQDEIADIIMKEVGKGYKDAKEVVRTADFRITYIEEALHMHGES 120
QY 123 DEGGKMDASTGHKLAVIRROPVGIIVLAIPNYPNLSGSKIAPALIGNVVMFKPTQOG 182
DB 121 MWGDSFPGGTSKLAIIORAPLGVLAIAFPNYPNLSAAKLAPALINGNAVIFKPAATQ 180
QY 183 SVSGVLAKAFEAAGLPGAVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGORIGKLA 242
DB 181 AISGIRKVAUHLKAGLPGVNVATGRGSAIGDYLVEHGINNVSTFGTNTGKLAKTA 240
QY 243 GMRPIMLELGGKDGAGIVLADADLNAAKQIVAGAYDYSQORCTAIKRVLVVEEVADELA 302
DB 241 AMPLVLELGGKDPGIVREDADLQDAANHIVSGAFYSQORCTAIKRVLVHENVADELVD 300
QY 303 KISNVAKLSVGDPPDNATVTPVIDNSADFIESLVDARQKGAKELEFEDGRLITPG 362
DB 301 LLKAQVAELSVGSPQDSTIVRLDDKSADFFVQGLVDVAKEGATIVIGNKRENLITYPT 360
QY 363 LFDHVTLDMLKAWEPFGPIPIIRVKDAEEAVATANKSDFGLOQSVFTRDFOKAFDIAN 422
DB 361 LIDHVTMVKAWEPFGPIPIIRVSSDEQALIANKSEFGLOQSVFTKDKAKAFAN 420
QY 423 KLEVGTVHNNKTRGPDPNFFPLGLKSGAGVQGIIRYSIBAMTNVKSIVLDMK 475
DB 421 KIETGSVQINGRTERGPDHPFFIGVKGSGWGAQGIKRSLSMTREKVTVLNFK 473

RESULT 10

US-10-369-493-17301
; Sequence 17301, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17301
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17301

Query Match 52.0%; Score 1244; DB 15; Length 481;
Best Local Similarity 51.4%; Pred. No. 1.5e-101;
Matches 240; Conservative 95; Mismatches 130; Indels 2; Gaps 1;

QY 10 NGEWKSS--VNQIEILSPIDSSILGFVPAMTREVVDHAKAGREALPAWALTYY 67
DB 15 NGEWERTGERISISAPASGVALGSPALSOEYVNDALQAKDAQKIWKPIHERVDL 74
QY 68 LHKAAADIERDKEEATVLAKESIKAYNASVTEVVRTADLIRYAAEGIRLSTSADEGK 127
DB 75 LYAWADLLEERKEIIGELIMHEVAKPKKSAIGEVSTADIRHTADEALRLNGETLKG 134
QY 128 MDASTGHKLAVIRROPVGIIVLAIPNYPNLSGSKIAPALIGNVVMFKPTQOGSGL 187
DB 135 FKSGSSKIALVEREPLGVLAISPYPNPNLAAKIALAPALVTGNTVWFKPATQSGS 194
QY 188 VLAKAFEAAGLPGAVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGORIGKLAGMRPI 247

Db 195 KMVEALADAGAPGIIQVVTGRTGSGVIGDHLVEHPGIDMITFTGGTTTGERISEKAKMPV 254
 QY 248 MLEGGKXAGIVLADADLNAAKQIVAGAYDYSQRCCTAKRVLVVEEVADELAEKISEN 307
 Db 255 VLEGGKXPAIVLDDADLKLASQIVSGAHSYSGRCCTAKRVPVQSVADQLVANIKEL 314
 QY 308 VAKLSVGGPDNDATVTPVIDNSADFTIESLVVDARQKAKELNBFKRDGRLLTPLGLFDHV 367
 Db 315 VEOITVSGPEDDADITPVIDEKSAFTQGLIDALENGATLLSGNKQGNLLSPITLDDV 374
 QY 368 TLDKMLAWEPFGPILPIIRVKDAEBAVAITAKSDFGLQSSVFRDFQKAFIANKLEVG 427
 Db 375 TPAMRVAREPFGPILPIIRVKDANEALISLNSQSYGLQASIFTKDTDRAINIKHLEVG 434
 QY 428 TVHNNKTGPDNPFPLGLKSGAGVQGIKYSIEMTNVKSIVLDM 474
 Db 435 TVHNAKTERGPDHFFPLGKSGLVGGIKPGLSLLSMTRERTVNL 481

RESULT 11

US-10-424-599-187423
 ; Sequence 187423, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 187423
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(503)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_140255C.1.pap
 US-10-424-599-187423

Query Match 43.4%; Score 1038.5; DB 12; Length 503;
 Best Local Similarity 46.2%; Pred. No. 2.7e-83;
 Matches 223; Conservative 80; Mismatches 163; Indels 17; Gaps 7;
 QY 5 YQYVNGWKSSVN--QIEILSPIDDSLSGFVPAMTRREVDHAMKAGREALPAWALTVY 62
 Db 17 KYIYADGVNKSAGKSVAILNPTTRKTQYKVCQSEEVNKMVMDLAKSAQKLWAKTPLW 76
 QY 63 ERAQYLHKAADIIRDEKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSTSA 122
 Db 77 KRAELHKAAILKEHKPTIAECLVKEIAKPAKDAVTEVRSGLVSVTAEEGVRI---L 133
 QY 123 DEGGKM--DASTGH---KLAIVTRQPGIVLAIAPYPNVNLSSKTAIPALIGNVVMEK 177
 Db 134 GEGKFLVSDSPFGNERTKYLTKSLPLGVIXAIPPPNYPNVLAVKTAIPALAGNSIVLK 193
 QY 178 PPTQGSVGLVIAKAFAEAGLPAGVNTITGRGSEIGDYIYEHEEVNFINTGTFPVQOR 237
 Db 194 PPTQGVASLHWHCFHLAGFPKGLINCVTGKSGEIGDFLTHWPGVNCISFTGG-DTGIA 252
 QY 238 IGLAGMRPIMLEGGKXAGIVLADADLNAAKQIVAGAYDYSQRCCTAKRVLVVEVA 297
 Db 253 ISKAGMPLQWELGGKXACIVLEISDLCKMKTSLYNKQPNPQRCCTAVKVLVWESAA 312
 QY 298 DELAEKISENVAKLSVGDPDFNATVTPVIDNSADFTIESLVVDARQKAKELNBFKRDGR 352
 Db 313 DALVEKVKAKVAKLTVGPPEDDCDITPVWSESSANFIEGLVLDAKEKVFDMGFACEX 372

QY 353 KRDRLLTGLFDHVTLDKMLAWEPFGPILPIIRVKDAEE-AVALANKSDFGLQSSVFT 411
 Db 373 KREGNPIWPLLLDXRPDIXIAWEEPGPVLPIRINSVEEIGHHCNANFGLKDMSPX 432
 QY 412 RPQKAFDIANKLEVGTVHNNKTGPDNPFPLGLKSGAGVQGIKYSIEMTNVKSIV 471
 Db 433 RDVNKAIMISDMETGTQINSAPARGDPHFPQGIKDSGISOQITNSINMMTKVKTIV 492
 QY 472 LDM 474
 Db 493 INL 495
 RESULT 12
 US-10-424-599-187419
 ; Sequence 187419, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 187419
 ; LENGTH: 434
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(434)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_140251C.1.pap
 US-10-424-599-187419

Query Match 41.9%; Score 1001.5; DB 12; Length 434;
 Best Local Similarity 48.6%; Pred. No. 4.2e-80;
 Matches 204; Conservative 73; Mismatches 132; Indels 11; Gaps 5;
 QY 5 YQYVNGWKSSVN--QIEILSPIDDSLSGFVPAMTRREVDHAMKAGREALPAWALTVY 62
 Db 17 KYIYADGVNKSAGKSVAILNPTTRKTQYKVCQSEEVNKMVMDLAKSAQKLWAKTPLW 76
 QY 63 ERAQYLHKAADIIRDEKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSTSA 122
 Db 77 KRAELHKAAILKEHKPTIAECLVKEIAKPAKDAVTEVRSGLVSVTAEEGVRI---L 133
 QY 123 DEGGKM--DASTGH---KLAIVTRQPGIVLAIAPYPNVNLSSKTAIPALIGNVVMEK 177
 Db 134 GEGKFLVSDSPFGNERTKYLTKSLPLGVIXAIPPPNYPNVLAVKTAIPALAGNSIVLK 193
 QY 178 PPTQGSVGLVIAKAFAEAGLPAGVNTITGRGSEIGDYIYEHEEVNFINTGTFPVQOR 237
 Db 194 PPTQGVASLHWHCFHLAGFPKGLINCVTGKSGEIGDFLTHWPGVNCISFTGG-DTGIS 252
 QY 238 IGLAGMRPIMLEGGKXAGIVLADADLNAAKQIVAGAYDYSQRCCTAKRVLVVEVA 297
 Db 253 ISKAGMPLQWELGGKXACIVLEISDLCKMKTSLYNKQPNPQRCCTAVKVLVWESVA 312
 QY 298 DELAEKISENVAKLSVGDPDFNATVTPVIDNSADFTIESLVVDARQKAKELNBFKRDGR 357
 Db 313 DALVEKVKAKVAKLTVGPPEDDCDITPVWSESSANFIEGLVLDAKEKVFDMGFACEX 372
 QY 358 LTPGLFDHVTLDKMLAWEPFGPILPIIRVKDAEBAVAITAKSDFGLQSSVFRDFQKA 417
 Db 373 LIWPLLDNVPDRIAWEEPGPVLPIKINSVEEIGHHCNANFGLQSCVCTKDVNKA 432

Db 236 HIAETISKISVLIEGKDPALVDDADFEIAANEIVKGAYGSGQRCTAKRVFVSHN 235
 QY 297 ADELAEKISENVAKLSVGDPDNATVTPVIDNSADFIESLVVDARQKGA--KELNBFKR 354
 Db 296 HOLLVNLINKVVDALTGVLGPQNPITPLINSLSKYNLSLVEDAIKGAIVHOKIVYNE 355
 QY 355 DGRILATGLEDHVTLDMKLAWEPFGFPIPIIRVKDAEAVAIAANKSDFGLOSSVFTEDF 414
 Db 356 KNNLHLPHIVDINTVTEMVAVBEPFGFPIPIITVNSIQBAIDLINQSGYGLQACIFTNY 415
 QY 415 QKAFDIANKLEVGTVHNNKTRGPNPFPFLGKSGAGVQGIIRYSYEAMTNVKSIVLD 473
 Db 416 ASTEQALQIESTGININKSSRGPPIIPFFGVKDSGFGVQIVDAILSWTTIKGLIIN 474
 RESULT 14
 US-10-268-518-4
 ; Sequence 4, Application US/10268518
 ; Publication No. US20030100034A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John Joseph
 ; TITLE OF INVENTION: 9136, A HUMAN ALDEHYDE DEHYDROGENASE
 ; FILE OF INVENTION: FAMILY MEMBER AND US\$S THEREFOR
 ; FILE REFERENCE: MP101-234P1RM
 ; CURRENT APPLICATION NUMBER: US/10/268,518
 ; CURRENT FILING DATE: 2002-10-10
 ; PRIOR APPLICATION NUMBER: 60/329,899
 ; PRIOR FILING DATE: 2001-10-16
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Consensus sequence
 US-10-268-518-4
 Query Match 32.9%; Score 786.5; DB 14; Length 492;
 Best Local Similarity 38.6%; Pred. No. 6,3e-61;
 Matches 193; Conservative 92; Mismatches 166; Indels 49; Gaps 16;
 QY 12 EWKSVN--QRIELSPIDDS--LGFVPMTRREVVDHAMKAGREAL---PAAALTVYERA 65
 Db 1 EWDVSAGKTFTEVPMANKGEVIGRVPENATIEDVDAAKAAKEAFKSGPWWAKVPASERA 60
 QY 66 OYLKADIIILERKEIATVLAKEISK--AYNASVTEVVRTADLIRYAA-----BEG 115
 Db 61 RIILKLADLEEREDELAALETJDLGKPLAEAKGDTVEVGRALDEIRYYAGWARKLMGER 120
 QY 116 I--RLSTSADGGKMDASTGHKLAVIRQPVGIVLAIPYNPVNLSGSKIAPALIGNV 173
 Db 121 VIPSILATDGE-----ELNYTRREPLGVGVISPWNPFLIALLKWLAPALAAAGT 170
 QY 174 VMFKEPQSGVSGSLVLAFAEAG---LPAGVFNITTGRSGEIGDYIVHEEVNFINTG 230
 Db 171 VIKRESQTPITALLLAELEEGAGNPPKGVNVPVPGFVGAEGVQALLSHPOIDKISFTG 230
 QY 231 STPVQQRIGKIA---GNRPIMLELGKQAGIVLADADLONAAKQIVAGAYDYSGQRCTAI 287
 Db 231 STEVGKLIMEAAAKNLKVTLELGKSPVIVFDADLDKAVERIVFGAFGNAGQVCIAP 290
 QY 288 KRVLVVEEVADELAKISENVAKLS--VGDPPDNAT--VTPVIDNSADFIESLVVDARQK 344
 Db 291 SKLLVHESIYDFEVLKERYKKLIGDPLDSDTNVYGLPISEQQDFRVLVSYIEDGKEE 350
 QY 345 GAK-----ELNEFKEDGRLLTPLGFDHVTLDMLKAWBEPGPIPIIRVKDAEAVA 396
 Db 351 GAKVLGGERDESKYLLGGVYVQPTITFDTPDMKIMKEIFGVLPIIKFDLDEATE 410
 QY 397 IANKSDFGLQSSVFTRD--FOCAFDIANKLEVGTVINNNKTRGPD--NFPFLGLK--GSGAG 453
 Db 411 LANDTEYGLAAVYFTKQILARAFRAKALEAGI WYNDVCVHAAPOLPGGVKQSGSIG 470

QY 454 VQ-GIRYSIEANTNVKSIVL 472

APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 1073
 ; LENGTH: 463
 ; TYPE: PRT
 ; ORGANISM: Methanococcus jannaschii
 US-10-369-493-1073

Query Match 31.7%; Score 759.5; DB 15; Length 463;
 Best Local Similarity 37.1%; Pred. No. 1.4e-58;
 Matches 175; Conservative 95; Mismatches 183; Indels 19; Gaps 8;
 QY 8 YVNGEWSVQ--IBILSPIDDSGLGFVPAMTRFEEVDHAMKAGREALPAAALTVYERA 65
 DB 2 FIDGK--INREDMVINPYSLEVIKIPALSREEAEIDTAEKYKVMKNUJITKRY 58
 QY 66 QYLKAAADIERKKEIATVLAKEISKAYNASVTEVVTADLIRYAAEGIRKLSADEG 125
 DB 59 NILAMIAKQIKKEKEELAKILDAIDAGKPIKQARVEVERSIGTFKLA--FYVKEHRDEV 115
 QY 126 GMDASTCHKLAVIRQPGVILALAPYNPVNLGSKITAPALGGNVVMEKPTQGSVS 185
 DB 116 IPSD----DKLITRREPVGIVGALTFFNPLNLGAHKAIAIATGIVNHVHPSPKAPLV 171
 QY 186 GLVLAK---AFEAAGLPAGVNTITGRGSEIGDYIVEHEEVNFINFTGTPVGQRIKGL 241
 DB 172 CIELAKIENALKKNVPLGVNLLTGAGEVGDVILVNEKVMISFTSSKVGSLITKK 231
 QY 242 AGMRPIMEIGGKAGIVLADLDNAKQIVAGAYDSGQRCATKRVLVVEEVADELA 301
 DB 232 AGFKKIALELGGVNPVILKADLNKAVNALIKGSPFIYAGQVCISVGMILVDESIADKFI 291
 QY 302 EXISENKAISVGDPPDNAT-VTPVIDNSADFTSLVVDAROKGAKELNEFKRGRLLT 360
 DB 292 EMFNKAKVUNGNPDEKTDVGLISVEHAWEKVEKAIDEGKLLGGKRDKALFY 351
 QY 361 PGLFPHVTLQMLAWPEEFGPIETIRVKDAEAAVAIAKSDFGLOSSVFTTRDFOKAPDI 420
 DB 352 PTILE-VORDNILCKTEFAPVPIRTNE-EEMIDIANSTEYGLHSAIFTNDINKSLKF 409
 QY 421 ANKLEVTGHINKTORGDPNPPFLGLKSGAGVQGIKRYSTAMTVKSVIL 472
 DB 410 AENLEFGGVINDSSLFRQDNMPFGVKGKGLGREGVKYAMEENSIKTI 461

RESULT 20
 US-10-369-493-1188
 ; Sequence 1188, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 1188
 ; LENGTH: 455
 ; TYPE: PRT

; ORGANISM: Methanobacterium thermoautotrophicum
 US-10-369-493-1188
 Query Match 31.0%; Score 741.5; DB 15; Length 455;
 Best Local Similarity 38.2%; Pred. No. 5.5e-57;
 Matches 173; Conservative 83; Mismatches 186; Indels 11; Gaps 5;
 QY 9 VNGEWSVQNIETILSPIDDSGLGFVPAMTRFEEVDHAMKAGREALPAAALTVYERAQYL 68
 DB 5 IDGEGVSGEKIFTVRNPFNGDEVDRVPLAGSDVERAIEAHRARDAMADLSARKISEKL 64
 QY 69 HKAADIIERDKKEIATVLAKEISKAYNASVTEVVTADLIRYAAEGIRKLSADEGGM 128
 DB 65 YVVADELKTELDEFARLITLESKPIRFSRDEKVSVEVARLSAEAGRL---YGESIPM 121
 QY 129 DASTGHK--LAVIRQPGVILALAPYNPVNLGSKITAPALGGNVVMEKPTQGSVS 186
 DB 122 DAGIGKGLTFTVRPLGVAAITFFNPLNLGAHKAIAIATGIVNHVHPSPKAPLSA 181
 QY 187 LVLAKAFAEAGLPAGVNTITGRGSEIGDYIVEHEEVNFINFTGTPVGQRIKGLAGMRP 246
 DB 182 LKWLILSE-HFPAGAVNAVTCGSEVGDVILDSPLVDKITTGSSVEVGRVYSARASMKK 240
 QY 247 IMEIGGKAGIVLADLDNAKQIVAGAYDSGQRCATKRVLVVEEVADELAKEKISE 306
 DB 241 ITTELGGNDPLIVMDDADIDSAAVAVRGSYLYSGQVCIAVKRMIVHEDVADEFADKLVN 300
 QY 307 NVAKLSVGDPPDNAT-VTPVIDNSADFTSLVVDAROKGAKELNEFKRGRLLT 365
 DB 301 ITGSRAGDPMDVTDVGPLINEDAAIEVERKVAEVEDGAEILCGSSRGNFVEFTVLD 360
 QY 366 HVTLDQMLAWPEEFGPIETIRVKDAEAAVAIAKSDFGLOSSVFTTRDFOKAFDIANKLE 425
 DB 361 HVVPGMEVIERETGVPSPPIRFGADEAIRANGICYALQAGVFTENIRTLARMARETE 420
 QY 426 VGTGHINKTORGDPNPPFLGLKSGAGVQGIK 458
 DB 421 AGTVLVNKQSTFRVDHMPF---GDSGAVENVR 449

RESULT 21
 US-09-738-626-6430
 ; Sequence 6430, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patentin ver. 3.0
 ; SEQ ID NO 6430
 ; LENGTH: 484
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6430

RESULT 25
US-09-815-242-10264
; Sequence 10264, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyakind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```

/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 10264
/ LENGTH: 462
/ TYPE: FRT
/ ORGANISM: Escherichia coli
/ US-09-815-242-10264

```

[illegible]

Query Match 30.3%; Score 726; DB 9; Length 482;
Best Local Similarity 34.3%; Pred.No. 1.4e-55;
Matches 164; Conservative 92; Mismatches 206; Indels 16; Gaps 7;

Qy 6 QNYNGEWKSSVN--QIELLSPIDDSLSLGFVPAMTREEVSHAMKAGREALPAWRAALTIVE 63

Dp 12 OALINGEWLDANNGEATVNTNPANGPKIGSVPKMGADETRAATDAANRALPAWRAALTAK 71

[illegible]

QY 124 EGGKNDASTGHKLAVIRRPVGLVALAPYNYPNVNLGGSKIAPALIGGNVWFKPPTQGS 183

Qy 184 VSGIVLAFAPAEACLPAGVFNTITGRGSEIGDYIVEHEEYVNFINTGSTPVGQIGKLAG 243

QY 244 --MRPIMUELGKDGAGIVLADADLDNAAKQIVAGAYDYSQRCFAIKRVLWVEVADELA 301

DD 247 KDJAKVSIHELGGINHAFFIVFDUDADLDRAVEGALASAEKNAQJTCVCANRDIUVQGVIDRFA 306
 QY 302 EKISENVAKLSVGDPFDN-ATVTPIDDNSADFIESLVDRQKGAKEI--NEFKRDR 357

Db 307 EKQAVSKLHIGDGLNGVTIGPLIDEKAVAKVEEHIALENGARVCGGKAHERGEN 386

Qy 358 LLTGLFQHVTVLDMKLAWEPGPILPIIRVKDAEEAVATANKSDFGLOSSVFTTRDFOKA 417

Db 367 FFQPTILVDVPANAKVSKETFGPLAPLFRFKDEADVIAQANDTEFGLAAYFYARDLSRV 426

Db 367 FQQTLLVDVFNARKVSKETEPGLAPLFRFKDEADVIAQANDTEFLAAYFARDLSRV 426
 QY 418 FDIANKLEVGTVHNNKTG-RGPNDFPLGLKSGAGVQGIYRSIEAMTNVKSIVLDM 474
 Db 427 FVYGEALBYGVGIN--TGLISNEVAPGGIKASGLGRGSGKYGIEDLYLEIKTMCIGL 482

RESULT 28
 US-10-282-122A-46932
 ; Sequence 46932, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 46932
 ; LENGTH: 483
 ; TYPE: PRT
 ; ORGANISM: Bacillus anthracis
 US-10-282-122A-46932

Query Match 30.2%; Score 721.5; DB 12; Length 483;
 Best Local Similarity 35.7%; Pred No. 3.6e-55;
 Matches 171; Conservative 82; Mismatches 203; Indels 23; Gaps 8;

QY 8 YNKGWSSVNNQIEILSPIDSSGLGFVPMTRVEVDHAKAGREALPAAALTVYERAY 67
 Db 16 YINGEWITLQIQEIVNNPATKIEIFATVPKGGVTEAKQVDAAEAFKSWSKLTAADRAK 75
 QY 68 LHKADIITRKEEIAITVLAKELSKAYNASVTEVTDLIRYABEGIRLSTSADEGK 127
 Db 76 LKKWFTLLDENKEEIAAIMTKEQGPFAEALGEVNVANSFVEWYAEGRV-----YGM 130
 QY 128 MDASTGHKLAVTRQPGVGLVLAIPYVNVLSGSKIAPALIGNVVMKPEPTQGSVSL 187
 Db 131 IFASHPNRIUWKQPGVGNAAITPNFPAMITRKVPALAAAGCTAVVKPASQTLTAL 190

QY 188 VLAKFAEAGLPAGVFNITGRSGSIGDYIVHEBEVNFNFSTGTPVQORIGKLAG--MR 245
 Db 191 KLAELAHRADIPKGVINIVTGSAKAIADTWEDGRVRKVSFTGSTEIGKELMASAAQTKW 250
 QY 246 PIMLELGGKAGIVLADLDNAKQIVAGAYDSGORCTAIRVLVVEEVADELAEKIS 305
 Db 251 KVSLELGGHAPFIYMDADLDKAVEAVIGSKFRNAGQTCICTNRVFOVEEYFAFVEKPE 310
 QY 306 ENVAKLSVGDPF-DNATVTPVIDDINSADFIESLIVDAROKGAKEINEFKR---DGRLLT 360
 Db 311 KAVGQKVGDFGDTGTVGLPDENAVSKVQEHIEDAIOKGGTVLYGQKVAEDGHHMQ 370
 QY 361 P---GLFDHVTLDKMLAWBEPFGPIPIIRVKDAEBAVAIAANKSDFGQSSVFRDFOKA 417
 Db 371 PTVIGLANDTMLCNW---EETGPVAPVAKETVEEVIERANHTPYGLAAVIFPKDISA 427
 QY 418 FDIANKLEVTGTVHNNKTGRGPD--NFFPLGLKSGAGVQGIYRSIEAMTNVKSIVLDM 474
 Db 428 FOISEALEYGIIGLNDGL--PVAQAPFGGFKBSGIGREGGHFGIEEYLEIKYISLGL 483

RESULT 29
 US-10-282-122A-60838
 ; Sequence 60838, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 60838
 ; LENGTH: 488
 ; TYPE: PRT
 ; ORGANISM: Listeria monocytogenes
 US-10-282-122A-60838

Query Match 30.0%; Score 718; DB 12; Length 488;
 Best Local Similarity 33.8%; Pred. NO. 7.5e-55;

Db 249 RLHMLGNNAPAVIFEDADIDAAADATAGSLKYAQRCASVSRVLAHESVHDELVSRIID 308
 QY 306 ENVAKLSVGDFFD-NATVTPVDDNSADFTIESLVDAQKQKAKELNEFKR---DG-RLLT 360
 Db 309 DAWAEMSGDGLFDDTTLGLFELVSADQADWAEVLVDVDRGATVVRGGERHVEDGWHYYE 368
 QY 361 PGLFHDVHLDMLKLAWEPPGPIPLIIRVKDAEBAVAIAANKSDFGIQQSVFTRDFQKAFDI 420
 Db 369 PTLADVPDARIIVDEEQGPVCATVTTVDDEDAVTSANGSELALDAAVTADHDEMRV 428
 QY 421 ANKLEVGTVHNNKTGRGPDNPFPLGLKSGAGVQGIQSVIRYSIEAMTNVKSIVL 472
 Db 429 ABRVNAGAVRINGAPSGELGDPFGGNDASGIGREGIDSTHFEVREKSIIL 480

RESULT 32

US-10-282-122A-59544
 ; Sequence 59544, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.

FILE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; PRIOR FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 59544
 ; LENGTH: 482
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae

US-10-282-122A-59544

Query Match 29.7%; Score 711; DB 12; Length 482;
 Best Local Similarity 34.1%; Pred. No. 3.1e-54;
 Matches 163; Conservative 90; Mismatches 209; Indels 16; Gaps 7;
 QY 6 QNYVNGEKSSWN--QLEILSPIDDSSLSGFPVPMATREEDVDMKAGREALPAWALTVE 63
 Db 12 QALIDQWRAPNGDVIVTNFANGELGSPVPMGADRETREIAANRALPAWALTVE 71

QY 64 RAQYLHKAADIIRKKEEATVLAKEISKAYNASVTEVVTRADLIRYAAEEGIRLSTAD 123
 Db 72 RANILRRWFDLMWENQDDLARIMTLEQKFLAEAKGEISYAASFIEWFAEGSKIYGDIT 131
 QY 124 EGGMDASTGHKLAVIRRPQGVIGVIAIAPYNYPNVLSGKIAPALIGNVVMFKPPTQGS 183
 Db 132 PGHQAD-----KRLIVIKQIPGVTAITPWNFPAAITRKAGPALAAGCTVWLKPASQTP 186
 QY 184 VSGVLAKAFAPAGIAGVENVITITGSGSEIGDYIVEHEEVNFINFTGTPVGOIRIGKLAG 243
 Db 187 FSALALAEANRAGIPAGVENVITITGSGSEIGDYIVEHEEVNFINFTGTPVGOIRIGKLAG 246
 QY 244 --MRPMLGELGKQAGIVLADADLNNAAKQIVAGDYISGQRCITAKELVYVEEVADELA 301
 Db 247 KDIKKVSLGNGNAPFIVFDDADLDKAVEGALASKFRNAGQTCVCANRLYVQDGVYDREA 306
 QY 302 EKISENVAKLSVGDFF--DNATVTPVDDNSADFTIESLVDAQKQKAKELNEFK---RDGR 357
 Db 307 EKLQQAWEKLRIGDGLQDGVTTTGLIDEKAVAKVEEHIADAIKAGKVVTTGKPHALGN 366
 QY 358 LLTPGLFDHVTLDMLKLAWEPPGPIPLIIRVKDAEBAVAIAANKSDFGIQQSVFTRDFQKA 417
 Db 367 FFOPTILNVNPDSSAKVAKEETFGPLAPLFRFDEADVIAQANDTEFGLAAYFYARDLSRV 426
 QY 418 FDIANKLEVGTVHNNKTG-RGPDNPFPLGLKSGAGVQGIQSVIRYSIEAMTNVKSIVLDM 474
 Db 427 FRVGEALVYGIIGIN--TGIISTEVAPFGVYKASGLREGSKYGIEDYLEIKYWCIGL 482

RESULT 33

US-10-369-493-18711
 ; Sequence 18711, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Jiongwei
 ; APPLICANT: Hankle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 18711
 ; LENGTH: 489
 ; TYPE: PRT
 ; ORGANISM: Halobacterium sp. NRC-1
 ; US-10-369-493-18711

Query Match 29.7%; Score 710.5; DB 15; Length 489;
 Best Local Similarity 35.9%; Pred. No. 3.5e-54;
 Matches 167; Conservative 84; Mismatches 199; Indels 15; Gaps 6;

QY 8 YVNGEW--KSSVNOILEILSPIDDSSLSGFPVPMATREEDVDMKAGREALPAWALTVE 65
 Db 15 YIDGEPVPPGRGTFLEVDDESTRELFTEIPAGTTDDVDAAYAAADAQTEWADTPQERA 74
 QY 66 QYLHKAADIIRKKEEATVLAKEISKAYNASVTEVVTRADLIRYAAEEGIRLSTAD 125
 Db 75 RIVAAASEILASHRDEVDVLLATESGTTTKGAAEFASVGTDEAASFTRWN-----G 129
 QY 126 GKMDASTGHKLAVIRRPQGVIGVIAIAPYNYPNVLSGKIAPALIGNVVMFKPPTQGSYS 185
 Db 130 DHKQSVVEKENIVRBPQGVIGIISPNWFFLNISIRAVAPAAAGNSVWLKPSANTPT 189
 QY 186 -GLVLAKAFAPAGIAGVENVITITGSGSEIGDYIVEHEEVNFINFTGTPVGOIRIGKLAG 244
 Db 190 GLLIAKFEAAGLPEGVNVNVTGKSEIGRIAGHPHADVISFTGTAIGRQVAATAGE 249


```

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8741
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Ralstonia metallidurans
; US-10-369-493-8741

Query Match      29.0%; Score 694; DB 15; Length 477;
Best Local Similarity 35.5%; Pred. No. 9.8e-53;
Matches 173; Conservative 83; Mismatches 197; Indels 34; Gaps 10;

QY 5 YQNVNGEWSVNOIBELSPIDSSLGFPVPMTRREEVD-HAMKAGREALPA----- 55
Db 3 YRNVIDGEMCDASG-----RTVDNVN-----PADTRDIVSRHAASDARDAAAAAFAAF 54

QY 56 --WAALTIVYERAOYLHKAADIIRDKKEIATVLAKEISKAYNASVTEVVRTADLIRYAAE 113
Db 55 DQWKKTTIGKRAKLINDAANLEAHDATIELTRECKALNLRDEWRSQATRFYAV 114

QY 114 EGIRLSTADEGGKMDASTGHKLAVIRROPVGVILAIAPYNPVNLGSGKIAPALIGNV 173
Db 115 EGQTFSGESYPNDPDP-----MLVYSRLREPLGVTVISPNWFPVSIPIARKIAPALITGNT 169

QY 174 VNFPPPTQSGVGLVIAKAPABAGLPAGVNPITGRGSEIGDYIVEHEEWNFTGSGTP 233
Db 170 VVFKPSDAPLSGLYLAERAFRAGIPKGVNLFITGSAEAVGTPTVESREVRASVPTGTS 229

QY 234 VQORTIGLAWRP-IMLELGGKAGIVLADADLDAKQIVAGAYDSQORTATKRVLV 292
Db 230 AGEQIHKSVPMTTRTQMEIGGRNPLIWNEDADLDRAVLAVKGLSLSGQACTGTGTRILV 289

QY 293 VEEVADELAEKISENVAKLSVGDPDFD-NATVTVPVDDNSADFIPLSVDAKQ-----GA 346
Db 290 MEDVQAFTEKLVAKYKALKIGSGWTAGMDLGLPLATKQLETVLSYIAIGKQEAATLLCGG 349

QY 347 KEINEFK-RGRLLITFGLPDHVTLDMLAWBEPFGPIIPILRVKDAEEAVATANKSDGL 405
Db 350 EQLSDGDFAHGYTVAPTVFTDTQSMRIAREE:FGVLAIIIEVDSYADAIAQANDTEYGL 409

QY 406 QSSVPTDFOKAFDIANKLEVGTTHNNKTGRGPDNFPFLGLKSGAGV--QGIRYSIEA 463
Db 410 SNAIATNPRYMHDFTRDIESGTVKINRTTGNLNVNAPFGGLKESSTSTFRESGRAGLEF 469

QY 464 MTNPKSI 470
Db 470 YTOIKTV 476

RESULT 41
US-10-282-122A-69547
; Sequence 69547, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

```

```

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69547
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
; US-10-282-122A-69547

Query Match      29.0%; Score 693; DB 12; Length 480;
Best Local Similarity 33.8%; Pred. No. 1.2e-52;
Matches 161; Conservative 89; Mismatches 212; Indels 14; Gaps 7;

QY 6 QNVNGE--KSSVNOIEILSPIDSSLGFPVPMTRREEVDHAMKAGREALPAWALTVYE 63
Db 12 QAYINGQWLDADGGQSKVNNPATNEILGVTPKGAETRAIEADKALPANRALTAKE 71

QY 64 RAQYLHKAADIIRDKKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEBEIRLSTAD 123
Db 72 RGNKLRWFELMIENQDGLLMTLEQKPLAEAKGEITVYAAAFIEWFAEEAKRVYGDVI 131

QY 124 EGGKMDASTGHKLAVIRROPVGVILAIAPYNPVNLGSGKIAPALIGNVNMFKPPTQS 183
Db 132 PGHQPD-----KRLIVLKQPIGVTAITPMNFFAAMITRAGPALAGCTWVLKPASQTP 186

QY 184 VSLGLVIAKAPABAGLPAGVNPITGRGSEIGDYIVEHEEWNFTGTPVGQRIQK--L 241
Db 187 FSALALAEAEERAGIPAGVSVTVTSGAGDISGELTGNPIVRKLSFTGSGTEIGRLMAECA 246

QY 242 AGMRPIMLELGGKAGIVLADADLDAKQIVAGAYDSQORTATKRVLVVEEVADELA 301
Db 247 KDIKKVSLELGGNAPFVIEDDADLDKAVEGAMISKYRNNGQTCVCANRIYVQDGVYDAFA 306

QY 302 EKISENVAKLSVGDPDFD-NATVTVPVDDNSADFIPLSVDAKQKAKELNEFKR-DGRLL 359
Db 307 EKLVAVAGKLGKINGLEDGITTGPLIDEKAVAKVKHIAADAVSKGATVLTGNSLSGSEFF 366

QY 360 TPLGLFDHVTLDMLAWBEPFGPIIPILRVKDAEEAVATANKSDFGLOSSVFTDFOKAFD 419
Db 367 EPTILNVSKDAAVAREETFGPLAFLRFKDEAEATALANDTEFGLASIFYAQNMSRVRP 426

QY 420 IANKLEVGTTHNNKTGRGPDNFPFLGLKSGAGVQGIIRYSIEANTVNSIVLDM 474
Db 427 VAEALEYGMVGIN--TGLISNELAPFGGKSGLREGSKYGIEDYLEIKYLCISV 480

```

RESULT 42

302 EKISENVAKLSVDPFF-DNATVTPVIDDSDAFESLWVDAROKGAKELNEFFKR-DGRLL 359
 307 EKLVAVGKLGKNGLEDGITTGLIDERKAVKVEHIAVSKGATVLTGNSLEGSPF 366
 360 TPLGFDHVTLDMLKAWERPFQPLPIRVKDAEVAJANKSDFGLQSSVFTTRDFQKAFD 419
 367 EFTLVNVSKDAVAREETFGPLAPLFRFKDEAEALALANDTBEGLASYFYAQNMSRVER 426
 420 IANKLEVTGTHNNKTGRGPNF-PFLGLKSGAGVCGIRYSIEAMTNVKSIVLDM 474
 427 VAALEYGMVGIN--TGLISNELAPFGGKSGREGSKYGIEDVLEIKYLCISV 480

RESULT 43

US-10-282-122A-75217
 ; Sequence 75217, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 75217
 LENGTH: 480
 TYPE: PRT
 ORGANISM: Salmonella typhimurium

US-10-282-122A-75217

Query Match 28.9%; Score 692; DB 12; Length 482;
 Best Local Similarity 33.3%; Pred. No. 1.5e-52;
 Matches 159; Conservative 95; Mismatches 208; Indels 16; Gaps 7;
 6 QNYNGEWKSS--VNGIETLSPIDSSLGFPVPMTRFEEVDHAKGREALPAWALTVE 63
 12 QAFIDGWRDARGDVPVSPANGKPLGNVPMKGAETTRDAINAANRALPALTAKE 71
 64 RAQYLKAAADIIRDKKEIATVLAKEISKAYNASVTEVTRADLIRYAAEGIRLSTSD 123

US-10-282-122A-70076
 ; Sequence 70076, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 70076
 LENGTH: 480
 TYPE: PRT
 ORGANISM: Pseudomonas syringae

US-10-282-122A-70076

Query Match 29.0%; Score 693; DB 12; Length 480;
 Best Local Similarity 33.8%; Pred. No. 1.2e-53;
 Matches 161; Conservative 89; Mismatches 212; Indels 14; Gaps 7;

6 QNYNGEW--KSSVNGIETLSPIDSSLGFPVPMTRFEEVDHAKGREALPAWALTVE 63
 12 QAYNGWLDAGGQSKVNNPATNEILGTVPKMGAAETTRAEAAADKALPAAALTAKE 71
 64 RAQYLKAAADIIRDKKEIATVLAKEISKAYNASVTEVTRADLIRYAAEGIRLSTSD 123
 72 RGNKLRWFELMIENQDGLMTLEQGRPLAEAKGKITVAASFIEWFAEAKRYVDYI 131
 124 EGGKWDASTGHKLVIRPQGVILAIAPYVNVLSGSKIAPALIGNVVMFKPFTQSS 183
 132 PQHQPD-----KRLVLKQPLGVTAALTWNFPAMITRKAGPALAAGCTWVLPASQTP 186
 184 VSGVLAKAFAPAGVAGVNTITGRSGBIGYIVVEHEVNFNFTGSTPTVQRIK--L 241
 187 FSALALAEALERAGIPAGVSVWTSAGDIGSELGTGNPIVRKLSFTGSTIEGRQMAECA 246
 242 AGMRPIMLEIGGKAGIVLADLDNAAKQIVAGDYSGORCTAIKRVLVVEEVADELA 301
 247 KDIKKVSLELGGNAPITVDDADLDKAVEGAMISKYRNNGQTCVCANRIYQDGYDAFA 306

Db 246 KDKKVSLELGGNAPFIVDDADLDKAVEGAILSKYNNNGQTCVCANRLYQDSVYDAFA 305
 QY 302 EKISENVAKLSVGGPDNDATVT--PVDDNSADFIESLVVDARQKAKELNEFK-RGRLL 359
 Db 306 EKLKAVAKLKIAGLEEGTTGPLIDEKAVAKVQEHADAVSKATVLAGGKPMGNEFF 365
 QY 360 TPLGFHDVTLDMKLAWEPRGPILPIIRVKDAEAAVAIAKNSDFGLOSVFTRDFQKAFD 419
 Db 366 EPTILTNPVKKAAVAKETFGPLAFPFKDEADVIANSDNTERGLASYFYARDLGRVER 425
 QY 420 IANKLVGTVHNNKTGR--GPDNFFPLGLKSGAGVQGIKSYIEAMTNVKSIVL 472
 Db 426 VAALEYGVGVN--TGLISNEVAPFGGKASGLGREGSKYGIEDYLEIKYICL 477

RESULT 45
 US-10-282-122A-51224
 ; Sequence 51224, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-03-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 51224
 ; LENGTH: 489
 ; TYPE: PRT
 ; ORGANISM: Bordetella pertussis
 ; US-10-282-122A-51224

Query Match 28.8%; Score 690; DB 12; Length 489;
 Best Local Similarity 33.4%; Pred. No. 2.3e-52;
 Matches 161; Conservative 93; Mismatches 200; Indels 28; Gaps 8;
 QY 8 YVNGKSWKSVN--QIELSPIDSSLG-----FVPAMTREVVDHAKAGREALPAWALTVE 63
 Db 19 YIDGKWVGAGNPSI--FVDNFTGKTIISVPKIGRKEAEQIDAEEALPAWARTGKE 76

Db 72 RANILRRWFLMMHQDDLRLMTLEQKPLAEAKGEISYAASFIEWFAEKGRIYDGI 131
 QY 124 EGGKMDASTGHKAVIRQPGVIVATAPNYPNLSKSTAPALIGGNVWFKPPTGS 183
 Db 132 PGHQAD-----KRLLVKIQIGVTAATTPWNPFSAITRKAGPALAAGCTWIKPEASGP 186
 QY 184 VSLVLAKAFABEAGLPAQVFNITGRGSEIGDYIVHEEVNFNFTGSTPVGQRIGKLAG 243
 Db 187 FSALAEALAQRAQIPAGVFNFTGSGAGDIGGELISNPLVKLSFTGSTEIGRLMEQCA 246
 QY 244 --MRPIMLELGGKAGIVLADLNAAKQIVAGAYDYGQRCTAIVKVLWVEEVADELA 301
 Db 247 KDKKVSLELGGNAPFIVDDADLDKAVEGAILSKYNNNGQTCVCANRLYQDSVYDAFA 306
 QY 302 EKISENVAKLSVGGPDNDATVT--PVDDNSADFIESLVVDARQKAKEL---NEEKRDGR 357
 Db 307 EKLKAVAKLKIAGLEEGTTGPLIDEKAVAKVQEHADALEXGARVITGGFAHKLGN 366
 QY 358 LTPGLFDHVTDLMLKLAWEPRGPILPIIRVKDAEAAVAIAKNSDFGLOSVFTRDFOKA 417
 Db 367 FQPTILADVDPNNAKAKETFGPLAFPFKDEADVIRQANDTFEGLAAVFFYARDLSRV 426
 QY 418 FDIANKLVGTVHNNKTGR--GPDNFFPLGLKSGAGVQGIKSYIEAMTNVKSIVLDM 474
 Db 427 FRYGALEYGVGVN--TGLISNEVAPFGGKASGLGREGSKYGIEDYLEIKYICGL 482

RESULT 44
 US-10-369-493-13957
 ; Sequence 13957, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 13957
 ; LENGTH: 478
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas fluorescens
 ; US-10-369-493-13957

Query Match 28.9%; Score 691; DB 15; Length 478;
 Best Local Similarity 33.5%; Pred. No. 1.8e-52;
 Matches 159; Conservative 92; Mismatches 209; Indels 14; Gaps 7;
 QY 6 QNYVNGKSWKSVN--QIELSPIDSSLGFPVAMTREVVDHAKAGREALPAWALTVE 63
 Db 11 QAFIDGAWVDADNGQTIKVNPAATGEILGVTFVFGAATRAIRAEADKALPAWALTAK 70
 QY 64 RAQYLHKAADIIBERDKKEIATVLAKEISKAYNASVTEVVRPADLIRYAAEEGIRLSTAD 123
 Db 71 RAKLRWFELMTENQDDLRLMTLEQKPLAEAKGEISYAASFIEWFAEAKRIYGVNI 130
 QY 124 EGGKMDASTGHKAVIRQPGVIVATAPNYPNLSKSTAPALIGGNVWFKPPTGS 183
 Db 131 PGHQPD-----KRLLVKIQIGVTAATTPWNPFSAITRKAGPALAAGCTWIKPEASGP 186
 QY 184 VSLVLAKAFABEAGLPAQVFNITGRGSEIGDYIVHEEVNFNFTGSTPVGQRIGK--L 241
 Db 186 FSALAEALAQRAQIPAGVFNFTGSGAGDIGGELISNPLVKLSFTGSTEIGRLMSCA 245
 QY 242 AGMRPIMLELGGKAGIVLADLNAAKQIVAGAYDYGQRCTAIVKVLWVEEVADELA 301

[illegible][illegible]

[illegible]

```

RESULT 48
US-10-369-493-4282
; Sequence 4282, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4282
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-369-493-4282

```

Query Match	20.5%;	Score 682.5;	DB 15;	Length 458;	
Best Local Similarity	36.1%;	Pred. No. 9.6e-52;			
Matches 168;	Conservative	84;	Mismatches 199;	Indels 15;	Gaps 7;
Qy	11	GEWKSVNQIELLSPIDDSGLGFPAMTREEDVHAKAGREALPAWALTYVERAQLHK 70			
Db	1	GERVTRPTLVDFFPYGTRVGTVELAS---VDVRAAFEAAYQQLTYERSQILER 57			
Qy	71	AADIIBERKEBIAVLAKIISKAYNASVTEVVRATDLIRVAABEGIRISTGADEGGMDA 130			
Db	58	AAAIMRTEREASDLISLESGLSKQDSRYELGFEVADVLKPAFIEALR---DDGQSFSC 112			
Qy	131	ST---GHKIAVI-BRQPV-GIVLAIAPNYPNVLGSGKIAPALIGNVMVKFPPTQGSV 185			
Db	113	DLTPHGKKRVRFSQREELAGVIAITFPENPMQVAHKIAPAIATNNRVLKPKSEKVPUS 172			
Qy	186	GLUIAKAFNAGLPAQVFNFTTCRSGEIGHYIIVEHEENVFNFTGSTPTVGQRIKGLAGMR 245			
Db	173	ALYIADVLYEAGLPAMLQLVTCDFRETADELITHPLAELWTFGGVAIGKTVIARAAYR 232			
Qy	246	PIMLEGGKADAGIADADLDNAAQIIVAGAYDSQCCTAIKRLVLWEVBADELAEKIS 305			
Db	233	RVLEJGGGNDPLIULDADUERATLVAQGSYKNSQCCTAVKMLYQKSVAAFDTLIV 292			
Qy	306	ENVAKLSVGDPDP-NATVTPVIDDNSADFIETSLVDARQKGAKELNEFKDGRLLTGPFL 364			
Db	293	EKTPAMTFGDPDPDASSQMGVTVIDVAAAQLFEARVNEAVASGARLLTGNQRNGALVAPTVL 352			
Qy	365	DHVLTDMLKIAWEEPPGILPIIRVKDAEEVAJANKSPDFGLQSSVTFTRDFQAFDIANKL 424			
Db	353	DGVDPSMTLVBREETFGPVSIIITFTDLDAIRISNGTAFGLSGSGCTNRQDAITRFINEL 412			
Qy	425	EVGVTHNNKTKGRGDNFPFLGLKSGAGV-QGIRYSIEAMTNVKS 469			
Db	413	RVGVNVMVEPYRIELTTPFGGKSGGSGYKEGVQAMKSFNLTK 456			

```

RESULT 49
US-10-369-493-7037
; Sequence 7037, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7037
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-369-493-7037

Query Match      28.5%; Score 682.5; DB 15; Length 458;
Best Local Similarity 36.1%; Pred. No. 9.6e-52;
Matches 168; Conservative 84; Mismatches 199; Indels 15; Gaps 7;

QY 11 GEMKSSVNOQIETLSPDSSILGFVPVPMTRREEVDHAMKAGSEALPAWAALTYYERAQVILHK 70
DB 1 GERVTRPRDLVDFDYTGTVGVPLAS---VDDVRAAFEYAAAQAQLTRYSQILR 57
QY 71 AADIIERDEKEIATVLAKEISKAYNASVTEVVRADLIKYAAEGIRLSTADESGKMDA 130
DB 58 AASIMRTERTEASDLISLSEGLSKQDSRYETIGRVADVIKFASTEALR-----DQQSFS 112
QY 131 ST---GHKLAVI-RRQPY-GVILAIAPYNPVNLSSGSKIAPALIGNVVMFKPTQGSVS 185
DB 113 DLTPHGKKRVSQREPLAGVIVAITFPHPNQVAKIAPATANNRVLLKPSEKVPLS 172
QY 186 GLVIAKAFAEAGLPAGVVENTITQGSBEIGDYIVHEEVEVNFINTGSTPVGQRIKGLACMR 245
DB 173 ALXLADVLYEAGLPAPMLQVLTGDPREIADLTHPLAEVLTVTTGTVAGTKYIARAAYR 232
QY 246 PIMLEIGGKQAGIVLADADLDNAKQIVAGADYSSQRCCTAIRXVLVVEEVADELAEKIS 305
DB 233 RVVLELGGNDPLIVLDDADUERAATLAVQGSYNSGQRCCTAVKRMVLVQKSVADFTDLVV 292
QY 306 ENVAKLSVGDPFD-NATVTPVIDNSADFTESLVVDQARQKAKELNEFKRDLGLTLPGLF 364
DB 293 EKTRAWTFGDPFDASSQMGTVIDVAQAQLFEARVNEAVASGARLLTGNQRNGALYAPTVL 352
QY 365 DHVTLDMKLAWBEPGPILPIIRVKQDAEBAVIAKNSDFGLQSSVTRFDQKAFIANKL 424
DB 353 GDVDPSTLVIREETFGVPSPITFTDLDDAIRISNGTAFGLSSGVCVTNQDAITFFINEL 412
QY 425 EGVGTVHNKTKRGDPNPPFLGLKGSAGV-QGIRYSITEAMTVNKS 469
DB 413 RVGTVMVWEVPGVRIELTPFGGIKDGLGYKEGVQEAAMKSFTNLKT 458

RESULT 50
US-10-369-493-17279
; Sequence 17279, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```

RESULT 50
US-10-369-493-17279
; Sequence 17279, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

Search completed: April 13, 2004, 14:28:21
Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:20:08 ; Search time 22 seconds
(without alignments)
1114.652 Million cell updates/sec

Title: US-09-868-195-12

Perfect score: 2393

Sequence: 1 LTKYQNVNGWSSVYQI.....GIRYSIEMTNVKSIVLDMK 475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/2/iaa/5A COMB.pap.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pap.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pap.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pap.*
5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pap.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	786.5	32.9	492	4	US-09-717-926-4	Sequence 4, Appli
2	775.5	32.4	492	4	US-09-634-955B-13	Sequence 13, Appl
3	733.5	30.7	526	4	US-09-252-991A-27302	Sequence 27302, A
4	711	29.7	487	4	US-09-489-039A-12633	Sequence 12633, A
5	705.5	29.5	491	4	US-09-585-174-3	Sequence 3, Appli
6	672	28.1	474	4	US-09-732-615-12	Sequence 12, Appl
7	668	27.9	594	4	US-09-252-991A-27749	Sequence 27749, A
8	667.5	27.9	481	4	US-08-976-063B-28	Sequence 28, Appl
9	659.5	27.6	508	4	US-08-956-171B-5241	Sequence 5241, Ap
10	658	27.5	486	4	US-09-328-352-4767	Sequence 4767, Ap
11	645	27.0	521	3	US-09-221-294-2	Sequence 2, Appli
12	642.5	26.8	507	4	US-09-489-039A-12722	Sequence 12722, A
13	635	26.5	498	4	US-09-489-039A-7569	Sequence 7569, Ap
14	632.5	26.4	505	4	US-09-543-681A-7151	Sequence 7151, Ap
15	632	26.4	496	4	US-09-543-681A-5701	Sequence 5701, Ap
16	631	26.4	529	4	US-09-252-991A-23785	Sequence 23785, A
17	628.5	26.2	505	4	US-09-328-352-5823	Sequence 5823, A
18	626	26.2	502	4	US-09-540-236-3780	Sequence 3780, Ap
19	625.5	26.1	510	4	US-09-252-991A-17138	Sequence 17138, A
20	624	26.1	510	4	US-09-134-001C-4541	Sequence 4541, Ap
21	623	26.0	493	4	US-09-328-352-6753	Sequence 6753, Ap
22	621	26.0	485	4	US-09-651-941-9	Sequence 9, Appli
23	621	26.0	485	4	US-09-955-597-9	Sequence 9, Appli
24	621	26.0	508	4	US-09-655-270A-9	Sequence 9, Appli
25	618	25.8	502	4	US-09-328-352-6687	Sequence 6687, Ap
26	616	25.7	482	4	US-09-155-183-4	Sequence 4, Appli
27	616	25.7	482	4	US-09-733-383-4	Sequence 4, Appli

28	609	25.4	497	1	US-08-513-841-2	Sequence 2, Appli
29	609	25.4	497	2	US-08-696-834-2	Sequence 2, Appli
30	609	25.4	497	3	US-08-942-673-2	Sequence 2, Appli
31	609	25.4	497	3	US-09-118-317-2	Sequence 2, Appli
32	607.5	25.4	518	4	US-09-134-001C-4451	Sequence 4451, Ap
33	604.5	25.3	498	4	US-09-328-352-5847	Sequence 5847, Ap
34	602	25.2	518	4	US-09-976-594-287	Sequence 287, App
35	601	25.1	506	4	US-09-134-001C-4383	Sequence 4383, Ap
36	600.5	25.1	514	4	US-09-489-039A-13820	Sequence 13820, A
37	598	25.0	490	4	US-09-328-352-6118	Sequence 6118, Ap
38	597.5	25.0	923	4	US-09-717-926-2	Sequence 2, Appli
39	594	24.8	505	4	US-09-328-352-7337	Sequence 7337, Ap
40	591.5	24.7	748	4	US-09-252-991A-28449	Sequence 28449, A
41	590.5	24.7	500	4	US-09-328-352-6757	Sequence 6757, Ap
42	585	24.4	499	4	US-09-543-681A-4515	Sequence 4515, Ap
43	583.5	24.4	488	4	US-09-134-001C-4246	Sequence 4246, Ap
44	581.5	24.3	492	4	US-09-489-039A-9023	Sequence 9023, Ap
45	578	24.2	502	4	US-09-489-039A-13971	Sequence 13971, A
46	577.5	24.1	481	4	US-09-328-352-4475	Sequence 4475, Ap
47	562.5	23.5	486	4	US-09-716-865-6	Sequence 6, Appli
48	562	23.5	475	4	US-09-252-991A-32298	Sequence 32298, A
49	561	23.4	642	4	US-09-252-991A-25398	Sequence 25398, A
50	555	23.2	564	4	US-09-328-352-7905	Sequence 7905, Ap

ALIGNMENTS

RESULT 1
US-09-717-926-4
; Sequence 4, Application US/09717926
; Patent No. 6569657
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Cook, William J.
; TITLE OF INVENTION: 32140, A No. 6569657el Human Aldehyde
; TITLE OF INVENTION: Dehydrogenase
; FILE REFERENCE: 35800/205243
; CURRENT APPLICATION NUMBER: US/09/717,926
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: US 60/214,707
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aldehyde dehydrogenase Pfam consensus sequence
US-09-717-926-4
Query Match 32.9%; Score 786.5; DB 4; Length 492;
Best Local Similarity 38.6%; Pred. No. 3.1e-65;
Matches 193; Conservative 92; Mismatches 166; Indels 49; Gaps 16;
Qy 12 EWKSSVN--QIETLSPIDDS--LGFVPANTREEDYHAKAGREAL---PAKALTYVERA 65
Db 1 EWDSAGKTFEWNPNKGEVGRVPEATAEDVDAVKAKEAFKSGWAKVPASERA 60
Qy 66 QYLKKAADIIERKEETATVLAKEISK--AYNASVTEVVTADLRVAA-----EEG 115
Db 61 RIUKKADIIERREDELAALATLDLKGPIAEAKGDTVEGRAIDEIRYAGWARKMGERR 120
Qy 116 I--RLSTADEGKMDASTGHKLAVIRKQPVGIVLAIAPYNPVNLSSKSTAPALGNNV 173
Db 121 VIPSLATDGE-----ELNYTRREPLGVGVISPNWFFLLALWKLAPALAAAGT 170
Qy 174 VNFKPPQSSVSLVAKAFBAAG---IPAGVFNTITGRSEIGDVIIVEHEEVNFINFTG 230
Db 171 VVLKPSQPLTALLAELEAEAGANNLPGVNVNVPFGAEVQALLSHDIPDIKISFTG 230
Qy 231 STPVQGRIGKLA---GMPIMLELGKDGAGIVLADLDNAAKQIVAGAYDSQGRCTAI 287

Db 231 STEVGKLMIAAAAKNLKVTLELGGKSPVIVFDADLDKAVRIVFGAFNAGQVCIAP 290
QY 288 KRLVAVVEVADELAELKISENVAKLS-VGDPDNAT--VTPVDDNSADFIESLVVDARQ 344
Db 291 SRLLVHESIYDFEVEKLERVKKLIGDPLSDTNYGLPSEQQFVJUSYIEDGKEE 350
QY 345 GAK-----ELNEFKRDGRLLTGLFDHVTLDMLKLANEPPFGPILPIRVKDAEAVA 396
Db 351 GAKVLCGGERDESKYELGGGYVQPTFTFDVTPDMKIMKEEFGPVLPIIKFOLDDEAIE 410
QY 397 IANKSDFGLQSSVFRD-FOKAFDIANKLEVGTTHNNKTGRGPD-NPFFLGL-KGSAG 453
Db 411 LANDTEYGLAAVFTKDLARAFRAKALEAGIIVWVNDVCVHAAPQLPFGVQKSSGIG 470
QY 454 VQ-GIRYSIEAMTNVKSIVL 472
Db 471 REHGKYGLEEVTEIKVTI 490

RESULT 2
US-09-634-955B-13
; Sequence 13, Application US/09634955B
; Patent No. 6511834
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Cook, William James
; TITLE OF INVENTION: 321481, 25964, 21686, NOVEL HUMAN DEHYDROGENASE
; FILE REFERENCE: WNI-134
; CURRENT APPLICATION NUMBER: US/09/634,955B
; CURRENT FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 60/192,002
; PRIOR FILING DATE: 2000-03-24
; SOFTWARE: FastSeq for Windows Version 4.
; NUMBER OF SEQ ID NOS: 35
; SEQ ID NO 13
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Aldehyde dehydrogenase family domain
US-09-634-955B-13

Query Match 32.4%; Score 775.5; DB 4; Length 492;
Best Local Similarity 38.4%; Pred. No. 3.3e-64;
Matches 192; Conservative 92; Mismatches 167; Indels 49; Gaps 16;

QY 12 EWKSSVN--QIILSPIDSS-LGFVPAMTRREVDHAKAGREAL---PANAALTIVYERA 65
Db 1 EWDSASGKTFEVNPNKGEVGRVPEATEEDVDAVKAKEAFKSGFWAKVPASERA 60
QY 66 OYLKAAADIIRDKKEIATVLAKEISK--AYNASVTEVTRADLIRYAA-----EEG 115
Db 61 RILKLAADLIEREDELAALETLDLCKPLAEAKGDTGVGAIDEIFRYAGWARKLMGERR 120
QY 116 I--RLTSPADEGKNDASTGHKLAVTRPOGVILAIAPYNPVNLSSGKIAPALIGNV 173
Db 121 VIPSLATDGE-----ELNYTRREPLGVGVISPNWFFLLALWKLAPALAGNT 170
QY 174 VMFPPTQSSVSLVIAKAPAEAG---LPAGVNTITGRSEIGDYIVHEEVNFNFTG 230
Db 171 VVLKPSQETPLTALLALIEEAGANNLPKGVNVVFGFAEVGQALLSHPDKISFTG 230
QY 231 STPVGQIRIGKLA---GMRPIMLELGGKDGAGIVLADADLDNAKQIVAGAYDYSQORTAI 287
Db 231 STEVGKLMIAAAAKNLKVTLELGGKSPVIVFDADLDKAVRIVFGAFNAGQVCIAP 290
QY 288 KRLVAVVEVADELAELKISENVAKLS-VGDPDNAT--VTPVDDNSADFIESLVVDARQ 344
Db 291 SRLLVHESIYDFEVEKLERVKKLIGDPLSDTNYGLPSEQQFVJUSYIEDGKEE 350
QY 345 GAK-----ELNEFKRDGRLLTGLFDHVTLDMLKLANEPPFGPILPIRVKDAEAVA 396

Db 351 GAKVLCGGERDESKYELGGGYVQPTFTFDVTPDMKIMKEEFGPVLPIIKFOLDDEAIE 410
QY 397 IANKSDFGLQSSVFRD-FOKAFDIANKLEVGTTHNNKTGRGPD-NPFFLGL-KGSAG 453
Db 411 LANDTEYGLAAVFTKDLARAFRAKALEAGIIVWVNDVCVHAAPQLPFGVQKSSGIG 470
QY 454 VQ-GIRYSIEAMTNVKSIVL 472
Db 471 REHGKYGLEEVTEIKVTI 490

RESULT 3
US-09-252-991A-27302
; Sequence 27302, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27302
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27302

Query Match 30.7%; Score 733.5; DB 4; Length 526;
Best Local Similarity 36.8%; Pred. No. 3.3e-60;
Matches 172; Conservative 87; Mismatches 186; Indels 23; Gaps 7;

QY 11 GEWK--SSVNQIILSPIDSS-LGFVPAMTRREVDHAKAGREALPANAALTIVYERAQYL 68
Db 53 GEWRHGRAGRRLKVNPFDDGSLLEIEQADRDLDAAVAKAAEVQPAWALGFSARAVAL 112
QY 69 HKAAADIIRDKKEIATVLAKEISKAYNASVTEVTRADLIRYAAEESGIRLSTA----DE 124
Db 113 YKAVEVDFRHEEIVDWIIRSEGS-----TLKAEIHWGAARAITLESASPARVH 163
QY 125 GSKMDASTGHKLAVTRPOGVILAIAPYNPVNLSSGKIAPALIGNVVMPKPTQGSV 184
Db 164 GRIVESDPKGSRYVRSAGVGVISPNWFFLHTQRSIAPALAGNAVYVVKPASDTPV 223
QY 185 -SGLVLAKAPAEAGLPAGVNTITGRSEIGDYIVHEEVNFNFTGSTPVCORIGKLAG 243
Db 224 CGULLARIIFEEAGLPAGLFSVWVVGSGEIGDAFVEHPVGLVFTFTGSTPVGRTGRIAS 283
QY 244 ---MRPIMLELGGKDGAGIVLADADLDNAKQIVAGAYDYSQORTAIKRVLVVEVADE 299
Db 284 GGAHLKRVHLELGGNSPFFVLGDADLEQAVNAVFGKFLHGGQICWAINCIIVEDSLYDA 343
QY 300 LAEKISENVAKLSVGDG--FDNAVTPVDDNSADFIESLVVDARQKAKELNEFKPDGR 357
Db 344 FAARFVERVGLRVGDPQADTA-VGVIVNAQLGKLEKIRLARQEGAKPLYEGVDGQ 402
QY 358 ILTPGLFDHVTLDMLKLANEPPFGPILPIRVKDAEAVAANKSDFGLQSSVFRDFOKA 417
Db 403 LLAPHVGEVATMETARDETFGELVGLLRARDEAHALELANASEYGLSSAVFSKDLERA 462
QY 418 PDIAKLEVGTTHNNKTGRGPDNFPFLGKSGAGVQGIYRISTEAMT 465
Db 463 VRFARQLRAGWTHVNDIPWDEANAPFGSEKSGLGRNGDWAIEEFT 510

RESULT 4
US-09-489-039A-12633

```
; Sequence 12633, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12633
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12633

Query Match 29.7%; Score 711; DB 4; Length 487;
Best Local Similarity 34.1%; Pred. No. 3.8e-58;
Matches 163; Conservative 90; Mismatches 209; Indels 16; Gaps 7;

QY 6 QNYVNGWKSVM--QIEILSPIDSSLGFPVPMATREEDVHAKGREALPAAALTYVE 63
DB 17 QAMIDGWRDAPNGDVIAVTPNANGEOIGSVPKGAGDETRAEIAANRALPAWRLTAKE 76
QY 64 RAQYLKKAADIIRDEKEETATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLSTAD 123
DB 77 RANILRRWFOLMMENQDDJRLMTLEQKPLAEAKGEISYAASFIEMAEGRKIYGTI 136
QY 124 EGKMDASTGHKLAVIRROPVGIIVLAIAPYVNPVNLGSKIAPALIGNVNMFKEPPTQGS 183
DB 137 PGHQAD-----KLLVVKIPGIVTAITPNPFAAMITRKAGPALAGCTWVKPASTP 191
QY 184 VSLGLVAKAFAGAGLPAGVFNITGRSEITGDYIVVEHEVNFNFTGSPVQORIGKLAG 243
DB 192 FSALALAEANRAGIPAGVFNVTGSGAGVGGELTSPNPLVKLSFTGSTEIGRQLMEQA 251
QY 244 --MRPIMLEGGKDGAGIVLADLDNAKQIVAGAYDYSGORCTAIKRVLVVEEVADELA 301
DB 252 KOIKVSLGNGNAPFIVFDDADLDKAVEGALASKERNAGQTCVCANRLVQDGYDRFA 311
QY 302 EKISENVAKISVGDPP--DNATVTPVIDNSADFIESLVVDARQKAKELNEFK--RDGR 357
DB 312 EKLOQAVEKLRIQDGLQDGYTTPGLIDKAVAKVEHIDALAKGAKVVTGKPHALGN 371
QY 358 LTPGLFDHVTLDKLAWEPEFGPILPIIRVKDAEAEVAIAANKSDGLQSSVTRDFOKA 417
DB 372 FFOPTILVNPDSAKVAKEETGPLAPLFRKDEADVIAQAANDTEFLGAAIFYARDLSRV 431
QY 418 FOIANKLEGVTHVHNKKG--RQPDNFPFLGLKSGAGVQGIYRYSIEAMTVKSLVDM 474
DB 432 FRVGEALRYGIIGN--TGIISTEVAPPFGVKAGLREGSKYIGIEDYLEIKYMCIGL 487

RESULT 5
US-09-585-174-3
; Sequence 3, Application US/09585174
; Patent No. 6586229
; GENERAL INFORMATION:
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Cattermole, Monica
; APPLICANT: Gatenby, Anthony A.
; APPLICANT: Gibson, Katherine J.
; APPLICANT: Ramos-Gonzalez, Isabel
; APPLICANT: Ramos, Juan
; APPLICANT: Sariafiani, Sima
; TITLE OF INVENTION: Method for the production of p-Hydroxybenzoate in Species of
; FILE REFERENCE: BC1018 US NA
; CURRENT APPLICATION NUMBER: US/09/585,174
; CURRENT FILING DATE: 2000-06-01
; NUMBER OF SEQ ID NOS: 112
```

```
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Pseudomonas mendocina KR-1
US-09-585-174-3

Query Match 29.5%; Score 705.5; DB 4; Length 491;
Best Local Similarity 34.5%; Pred. No. 1.2e-57;
Matches 167; Conservative 94; Mismatches 198; Indels 25; Gaps 8;

QY 5 YONY-----VNGEWK--SSVNOQIEILSPIDSSLGFPVPMATREEDVHAKGREALPAA 58
DB 8 YQMLELOPLAGOWRAGSGRPLEVDFPNDELLRIALASREDLDAAYRKARQOREWAT 67
QY 59 LTVYERAOYLHKAADIIRDEKEETATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRL 118
DB 68 TAPARARVLEAVKIFDERREIIMIIRESGS-----TRIKAQIEWGAARAITL 118
QY 119 STSADG--GKMDAST--GHKLAVIRROPVGIIVLAIAPYVNPVNLGSKIAPALIGNVV 174
DB 119 ESASLPNRVHGRIITASNISGKSRVYRAPLGVIGVISPWNFPLHLTARSLAPALGNV 178
QY 175 MFKPTQGSVS--GLVLAKAFAGAGLPAGVFNITGRSEITGDYIVVEHEVNFNFTGSP 233
DB 179 VVKPASDPTITGGILLARIFEAGLPAGVLSVVVGSGAIEGDAFVEHPVALLISFTG 238
QY 234 VGORIGKLAG---MRPIMLEGGKDGAGIVLADLDNAKQIVAGAYDYSGORCTAIKR 289
DB 239 VGNIGRTASGGEHLKHALELGGNSPFVVLADADVEQAVNAVVGKFLHQGQICMAINR 298
QY 290 VLVVEEVADELAEKISNNAKLSVGDPPDNATVT--PVIDNSADFIESLVVDARQKAKE 348
DB 299 IIVEQPLEDDTRFRVVRVKALPYGDPSPGTVVGVINARQLAGLEKIATAKAGATL 358
QY 349 LNFEDKDGRLTPGLFDHVTLDKLAWEPEFGPILPIIRVKDAEAEVAIAANKSDGLQSS 408
DB 359 LLGGEFQGNVMPHFVGNVTADMETAREIEFGPLVIGLQSGARDAEHALELANSSEYGLSSA 418
QY 409 VTRFDQKAFDIANKLVGVTHVHNKKG--RQPDNFPFLGLKSGAGVQGIYRYSIEAMTVK 468
DB 419 VFTASLERGVQFARRIHAGTHVNDI--PVNDEPNAPFGGKNSGLGRFNGDWAIIEFTTDH 478
QY 469 SIVL 472
DB 479 WITL 482
```

```
RESULT 6
US-09-732-615-12
; Sequence 12, Application US/09732615
; Patent No. 6632650
; GENERAL INFORMATION:
; APPLICANT: Chen, Mario W.
; APPLICANT: Chen, Qiong
; APPLICANT: Gibson, Katherine J.
; APPLICANT: Kostichka, Kristy N.
; APPLICANT: Thomas, Stuart M.
; APPLICANT: Nagavajan, Vasantha
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 US NA
; CURRENT APPLICATION NUMBER: US/09/732,615
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/170,214
; PRIOR FILING DATE: December 10, 1999
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Rhodococcus ruber
US-09-732-615-12
```

Query Match 28.1%; Score 672; DB 4; Length 474;
 Best Local Similarity 35.5%; Pred. No. 1.7e-54;
 Matches 169; Conservative 84; Mismatches 197; Indels 26; Gaps 9;

QY 8 YNNGWKSQVNO--IEILSPIDDSGLGFVPAMTREVVDHAKGREALPAAALTYERA 65
 DB 10 YNNGWASTSTKTVIEVLNPAETEVIGTVPDGAADVDAVAARAFAFDGMASTPVDKRA 69
 QY 66 QYLHK-AADIIRKEEIAVLAKESKAYNASVTEVVTADLIRYAAEGIRLSTADE 124
 DB 70 QYLRAIAGIADRDELARTISAEMGAPLSFAQMOVPLPINSFSHAABAESFFPERTE 129
 QY 125 GGDADSTGCHKLAVIRROPVGIYLAIPYVNVNLSGSKIAPALIGNVWMEKPTQGSV 184
 DB 130 GS-----SVIVREPIGVGAIPTWNVLHQIAAKVAVALAAGNTIVVKPSEVAPL 179
 QY 185 SGLVLAKAFABAGLPAGVNTITGRSGEIGDYIVVEHEVNFINTGSTPVQORIGKLAG- 243
 DB 180 NAMLAELIIDAGVPASVFNLSGTPVVGEGALSHHEVDMISFTGSTNAGKRVSELAQA 239
 QY 244 -MRPIMLEGGKAGIVLADADLDNAKQIVAGAYDSQRCCTAKRKVLVVEVADELA 302
 DB 240 TVKRVALEGGKANSIIVLDDADIDELMPENAVQWAMINSQTCGSAITRLVPRALITE-AB 298
 QY 303 KISENVAK-LVSGPFD-NATVTPVIDNSADFTESLVDARQKA-----KELNEFKR 354
 DB 299 TAAKTAAEATVAGAPDDPDTLGLPVSATQKRVGVIDRGVQEGATLITGGSEPVGLA 358
 QY 355 DGRLLPGLFDHVTLDKMLAWEPFGPIPIIRVKDAEEAVAIANKSDFGLOQSVFTROF 414
 DB 359 VGYVXTFIFSEVPTDTHREIBFGVLSIAPYDTEADAVIANDSEYGLRGVWNRDV 418
 QY 415 OKAFDIANKLEVGTWHNNKTGR-PDNPFPLGLKSGAGVQGRISIEAMTVKSI 470
 DB 419 DRARAAVAMRTQWING--GEFNEFAPFGYKQSGTGREGFTGHEFFLEIKSL 472

RESULT 7
 US-09-252-991A-27749
 ; Sequence 27749, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 27749
 ; LENGTH: 594
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27749

Query Match 27.9%; Score 668; DB 4; Length 594;
 Best Local Similarity 33.2%; Pred. No. 5.6e-54;
 Matches 158; Conservative 90; Mismatches 212; Indels 16; Gaps 7;

QY 6 QNYVNGEWSVN--QIELSPIDDSGLGFVPAMTREVVDHAKGREALPAAALTYE 63
 DB 123 QAYVDGAWDADNGQITKNNPATGTEIGSVPKGAAETREAEADKALPAWALAKE 182
 QY 64 RAQYLHKAADIIRKEEIAVLAKESKAYNASVTEVVTADLIRYAAEGIRLSTAD 123
 DB 183 RANKURRWFDMITENQDILUMITEQKPLAEKGEIYASLEWFEAKRIYDGTI 242
 QY 124 EGGKMDASTGCHKLAVIRROPVGIYLAIPYVNVNLSGSKIAPALIGNVWMEKPTQGS 183

DB 243 PGHQPD-----KRIIVKQPIQVTAITPNNFSPAMITKAGPAAAGCTWVUKPASQTP 297
 QY 184 VSGVLAKAFABAGLPAGVNTITGRSGEIGDYIVVEHEVNFINTGSTPVQORIGK--L 241
 DB 298 YSALAELAEARAGIPAGVSVVTSAGVGGELTSNPVIRKLITGSGTIGRLQMAECA 357
 QY 242 AGMRPIMLEGGKAGIVLADADLDNAKQIVAGAYDSQRCCTAKRKVLVVEVADELA 301
 DB 358 QDIKKVSLLEGNAFFIVDDADLDAAVEGALISKYRNNQTCVCANRLYVQDGVDAFV 417
 QY 302 EKISENVAKLSVGPFD-NATVTPVIDNSADFTESLVDARQKAGKELNEFK---RDGR 357
 DB 418 DKLAAVAKLNTINGLSEAGVTTGPLIDAKAVAKVEEHADAVSKAKVVGSGPHALGT 477
 QY 358 LLTPGLFDHVTLDKMLAWEPFGPIPIIRVKDAEEAVAIANKSDFGLOQSVFTROFKA 417
 DB 478 FFPTLIVDPKNAIVKSDITFGLAPVRFKDEAEVIAVMSNDTEFGLASVYFADLARV 537
 QY 418 FDIANKLEVGTWHNNKTGR-PDNPFPLGLKSGAGVQGRISIEAMTVKSI 472
 DB 538 PRVAOLEYGMVGIN--TGLISNEVAPFGGKASGLREGSKYGIEDYLEIKYLC 591

RESULT 8
 US-08-976-063E-28
 ; Sequence 28, Application US/08976063E
 ; Patent No. 6524831
 ; GENERAL INFORMATION:
 ; APPLICANT: Steinhuchel, Alexander
 ; APPLICANT: Priefert, Horst
 ; APPLICANT: Rabenhorst, Jurgen
 ; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
 ; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
 ; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
 ; FILE REFERENCE: Bayer-9998-CAO
 ; CURRENT APPLICATION NUMBER: US/08/976,063E
 ; CURRENT FILING DATE: 1997-11-21
 ; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
 ; PRIOR FILING DATE: 1996-11-29
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 481
 ; TYPE: PRT
 ; ORGANISM: not required under old rule
 US-08-976-063E-28

Query Match 27.9%; Score 667.5; DB 4; Length 481;
 Best Local Similarity 36.4%; Pred. No. 4.5e-54;
 Matches 169; Conservative 80; Mismatches 202; Indels 13; Gaps 6;

QY 15 SSVNQIELSPIDDSGLGFVPAMTREVVDHAKGREALPAAALTYERAQYLHKAADI 74
 DB 16 SDERTFEERSPLTGEVSVRAAASLEDAADAAVAAQAAPPEWAAALAPSEERRARLRAADL 75
 QY 75 IERDKEEIAVLAKESKAYNASVTEVVTADLIRYAAEGIRLSTADEGKMDASTGH 134
 DB 76 LE-DRSSEFTAAASGTAGAGNMGVFNLYLAAGMLREAA-----AMTIOQGVVFSNVP 129
 QY 135 KLAVIRROPVGIYLAIPYVNVNLSGSKIAPALIGNVWMEKPTQGSVGLVLAKAFA 194
 DB 130 SFAMAVROPQGVIGLAPNPNPVLGVRVAVNPLACGNTVVLKSSSELPFTHRLIGQVLH 189
 QY 195 EAGLPAGVNTITGRSE---IGDYIVVEHEVNFINTGSTPVQORIGKLAG--MRPIML 249
 DB 190 DAGLGDGVNVNISNAQDAPAVVERLIANPVRNVNFTGSTHVGRIIGELSAHILKPAVL 249
 QY 250 ELGGKAGIVLADADLDNAKQIVAGAYDSQRCCTAKRKVLVVEVADELAEKISENVA 309
 DB 250 ELGGKAPFLVLDADLDAAVEAAAFAYFNQOQICNSTER-LIYTVADAFVEKLARKVA 308
 QY 310 KLSVGDGPDNATVI-PVIDNSADFTESLVDARQKAGKELNEFKRGLLTPGLFDHVT 369

Db 309 TLRAGDPQSVGLSLIDANAGRIQVLVDLALAKAGQVGGGLDGSIMQPMLLIDQVT 368
 Qy 369 LDKLWEPFGPILPIIRVKDABEAVATANKSDFGLQSVFTRDQKAFDIANKLEVGT 428
 Db 369 EEMLYREESFGFVAVLRGDEELRLRLANDSEGLSAAIFSRDVSRAAMELAQRVDSGL 428
 Qy 429 VHNKTKGRDPNFFFLGKSGAGVQGIIRYSIEAMTNVKSIVL 472
 Db 429 CHINGPTVDEAQMFGVKSXGSGFGSRAISIEHFTOLRLTI 472

RESULT 9

US-08-956-171E-5241
 ; Sequence 5241, Application US/08956171E
 ; Patent No. 6593114
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; Gil H. Choi
 ; Patrick S. Dillon
 ; Craig A. Rosen
 ; Steven C. Barash
 ; Michael R. Fannon
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 5256
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/956,171E
 Filing DATE: 20-Oct-1997
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 66/009,861
 Filing DATE: January 5, 1996
 APPLICATION NUMBER: 08/781,986
 Filing DATE: January 3, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Mark J. Hyman
 REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PB248P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224
 TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 5241:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 508 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 5241:
 US-08-956-171E-5241

Query Match 27.6%; Score 659.5; DB 4; Length 508;
 Best Local Similarity 34.0%; Pred. No. 2.8e-53;
 Matches 166; Conservative 99; Mismatches 198; Indels 25; Gaps 10;

Qy 1 LTKEQNYVNGEW--KSSVNIQLSPIDSSGLGFVPAMTREVVDHAKAGREALPAWAA 58
 Db 22 IAEYGLFNGFVNGSSDETIEVTPATGETLSHITRAKDQDVHAKVQAQAFESWSL 81
 Qy 59 LTVYRAQLHKAADIIRDKBEIATVLAKEISKAYNASVTEVTRADLIYAAEE-GTR 117
 Db 82 TSKSERAQMLRIGDKLMAQKDKIAMTETLNNKPIRET-----TAIDIPFAARHIFY 135

Qy 118 LSTADSGKMDASTGHKLAVIREQPVGIVLAIAPIYVNVNLSGSKIAPALIGNVVFK 177
 Db 136 ASVETEEGTVDIDKDTMSIVRHEPIGVGVAVANFPMLLAANKIAFAIAGNTIVIQ 195
 Qy 178 PTOQSGVSLVAKAFARAGLPAGVFNITGRSGEIGDYIVVEHEVNFNFTGSTPVGQR 237
 Db 196 PSSSTPLSLLEVAKIFQEV-LFKGVNLTGKSSGNAIFNHDGVDKLSFTGSTDVGYQ 254
 Qy 238 IGKLAG--MRPIMLELGGKAGIVLADADLDAAKQIVAGAYDYSQRCCTAKRVLVVEE 295
 Db 255 VAEEAAKHLVPATLELGKKSANIILDDANLDLAVEGILQILFNQGEVCSAGSRLLVHEK 314
 Qy 296 VADELAEKISENVAKLSVGDPDNAT-VTVVIDDSADFIESLVVDARQKAKEL---NE 351
 Db 315 IYDQLVPRLOEAFSNIKVGNPQDEATQMSQTGKQDLKIOSYIDAAKESDAQIILAGGHR 374
 Qy 352 FYRDRLLTFLGLFDHVTI-----DMKLAEEEPFGPILPIIRVKDABEAVAIANKSDFGL 405
 Db 375 LTENG--LDKGFPEPTLIAVDPNHHKLAQEEIFGPVLIVIKVDDQEAIDIANDEYGL 432
 Qy 406 QSVFTRDFOKAFDIANKLEVGVTHNNKTGRGDPNFFPLGLKSGAGVQGIIRYSIEAMT 465
 Db 433 AGGVFSQMITALNIAKAVRTGRIWINT-YNQVPEGAPFGYKSGIGRETYKGLSNYQ 491
 Qy 466 NVKSIIVLD 473
 Db 492 QVKNIVID 499

RESULT 10

US-09-328-352-4767
 ; Sequence 4767, Application US/09328352
 ; Patent No. 6562958

GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4767
 ; LENGTH: 486
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-4767

Query Match 27.5%; Score 658; DB 4; Length 486;
 Best Local Similarity 32.2%; Pred. No. 3.6e-53;
 Matches 154; Conservative 92; Mismatches 219; Indels 14; Gaps 6;

Qy 4 EYQNYVNGEWSVNVQ--IBILSPIDSSGLGFVPAMTREVVDHAKAGREALPAWAA 61
 Db 14 QQQAYINGQWLAQSNATVPVSNPATGEIGTIPNMGAAEATQAVEAAVYATALQSMKALTA 73
 Qy 62 YERAQYLHKAADIIRDKBEIATVLAKEISKAYNASVTEVTRADLIYAAEEGIRLSIS 121
 Db 74 QNRADILLAAHKLVLDDHTDELALIMTIEQCKPLAEAKGEVRYAASFQWFAEEGKRI--- 130
 Qy 122 ADEGKMDASTGHKLAVIRQPVGIVLAIAPIYVNVNLSGSKIAPALIGNVVNFKEPTQ 181
 Db 131 --YGVPIVNNQQRFIISKEPVGVAAITPNPFIAMITRKAAPALAAAGCTTVVVKANE 188
 Qy 182 GSVGLVLAFAAEAGLPAGVFNITGRSGEIGDYIVVEHEVNFNFTGSTPVGQRIGKL 241
 Db 189 TPYCALAIKLAKEAGIPAGVINVTGKSQELSGVFTSHEKVKVLTFTGSTPVGRLMQQ 248
 Qy 242 AG--MRPIMLELGGKAGIVLADADLDAAKQIVAGAYDYSQRCCTAKRVLVVEVADE 299
 Db 249 CSSTIKKLALELGGNAPLIVFDDADLDKAVQGAIFAKFRNAGQTCVCANRIYVHDIYQA 308
 Qy 300 LAEKISENVAKLSVGDPF-DNATVTVVIDDSADFIESLVVDARQKAKELNEFRDGG-- 356

Db 309 FAEKFEVQVQKFGVNGLEDCVQIGPLINEXAVLKAQQLIDDAVSKGAKIACGKQHAG 368
 QY 357 -RLLTPLGLFDHVTLDKMLAWEPFPIIPRIIRVKDAEAEVAIAKNSDFGLQSSVTRDFQ 415
 Db 369 QTFEVSVLTVNDRTEIMEIQEIEFGFVAPLIRFTDEADVAAQANDTIFGLAAVYVSENI 428
 QY 416 KAFDIANKLEVGTTHNNKTGRGPNFFFLGLKSGAGVQGIIRYSIEAMTNVKSIVLDM 474
 Db 429 RLNRVSEQLGYGMVGM-NATAISNEVPFGVQSGVREGSKYGLBEFTIKYMCGL 486

RESULT 11

US-09-221-294-2
 ; Sequence 2, Application US/09221294
 ; Patent No. 6268138
 ; GENERAL INFORMATION:
 ; APPLICANT: Riccardo Dalla-Favera and
 ; APPLICANT: Alessandro Massimo Gianni
 ; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
 ; TITLE OF INVENTION: Aldenhyde Dehydrogenase-1 Gene and Uses of Said
 ; TITLE OF INVENTION: Vector
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM 330 466 DX2
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.24
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/221,294
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 42990-B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-278-0400
 ; TELEFAX: 212-391-0525
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 521 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-221-294-2

Query Match 27.0%; Score 645; DB 3; Length 521;
 Best Local Similarity 34.1%; Pred. No. 6.6e-52;
 Matches 164; Conservative 85; Mismatches 210; Indels 22; Gaps 10;
 QY 8 YNGEWKSSVN--QIEILSPIDSSLGFPVPMTRREVVDHAKAGREAL---PAAALTVY 62
 Db 40 FINNEHDSVSGKKFPFNPATPEELQVBEEDKVDKAVKAARQAFQIGSPWMTDAS 99
 QY 63 ERAQYLKAAADIIRDKKEIATVLAKEI---SKAY-NASVTEVVRTADLRVAAEEGIRL 118
 Db 100 EGRLLYKLADLERD-LLATMESMSMGKLYSNAYLNDLAGICKILRVCAWADKI 158
 QY 119 STSADEGKMDASTGHKLAVIRQPGVGLVLAIPYNNPVLNSGSKIAPALIGNNVMPK 178
 Db 159 QCG---GRTPIDGNFFYTHREPIGVCGQIIPWNPFLWLIWKIGPALSGNTVVVVKP 214
 QY 179 PTQGSVSLVLAKAFAGAGLPGVNTITGRGSEIGDYIVHEEVEVNFNFTGSPVQRI 238
 Db 215 AEQTPLTALHVASLIKEAGFPFVNVIVPGVGTAGAAITSSHMDDIKVAFPTGSTGVKLI 274

QY 239 GKLAG---NRPMLELGGKDAIGIVLADADLONRAKQIVAGAYDYSQORCTALKVVLVVE 295
 Db 275 KEAAGKSNLKRVTLELGGKFCIVLADADLONAVEFAHGVFVHQGCCICIAASRIFVEES 334
 QY 296 VADELAEKISENVAKLSVGDPF-DNATVTPVIDDNSADFTESLWVDARQKAK---ELNE 351
 Db 335 IYDEFVRSVERANKYILGNPLTFPGVTQGPQIDREQVDKLLDLESKKEGKALECGGP 394
 QY 352 FKRGRLPLPGFLDHVTLDKMLAWEPFPIIPRIIRVKDAEAEVAIAKNSDFGLQSSVT 411
 Db 395 WNGKGVFVQTVFSNVTDENRIAKEELFGPVQIQIMKFKSLDDVIKRNNTFYGLSAGVT 454
 QY 412 RPFQAFDIANKLEVGTTHNNKTGRGPNFFFLGLKSGAGVQGIIRYSIEAMTNVKSIV 471
 Db 455 KQIDKAITISSALQAGTWV-NCYGVVSAQCFFGFGKMGNGELGEYGFHEHYTEVKTIV 513
 QY 472 L 472
 Db 514 V 514

RESULT 12
 US-09-489-039A-12722
 ; Sequence 12722, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 12722
 ; LENGTH: 507
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-12722

Query Match 26.8%; Score 642.5; DB 4; Length 507;
 Best Local Similarity 32.9%; Pred. No. 1.1e-51;
 Matches 162; Conservative 99; Mismatches 196; Indels 35; Gaps 13;
 QY 4 EYONYVNGKSSVN--QIEILSPIDSSLGFPVPMTRREVVDHAKAGREALPAAALTV 61
 Db 23 EQQLYNGGYTSATSGRTETINPATGEVLATVQAAGREDVDRAVESAQRGOKIWAAMTA 82
 QY 62 YERAQYLKAAADIIRDKKEIATVLAKEISKAYN-ASVTEVVRTADLRVAAEEGIRLST 120
 Db 83 MERSILRRVDDLIRQRNDELARLETDTGKPLSETAAVDIVTGTVDLEYA--GL---I 137
 QY 121 SADEGKM---DASTGHKLAVIRQPGVGLVLAIPYNNPVLNSGSKIAPALIGNNVMPK 177
 Db 138 PALEGSQPLRDSG----FVYTRREPQVGVAGICAWNPYIQLALWKSAPALAGNANIFK 193
 QY 178 PTQGSVSLVLAKAFAGAGLPGVNTITGRGSEIGDYIVHEEVEVNFNFTGSPVQQR 237
 Db 194 PSEVTPALAKAIIYSEAGLPGDVENVLPFGIAGETQGLTEHPDIAKISFTGGVASKK 253
 QY 238 I---GKLAMGPIMLELGGKDAIGIVLADADLONRAKQIVAGAYDYSQORCTALKVVLVVE 294
 Db 254 VMANSAASLSEVTELMELGGKSELIITADADLDAADIAAMNFISSQVCTGTGTVFPA 313
 QY 295 EVADELAKISENVAKLSVGDPF-DNATVTPVID---DNSADFTESLWVDARQKAKEL 349
 Db 314 KQAEFEKILLERVARIRAGDLFADDTNFGELVSPFHRDNLVRYES---GKRGARLL 369
 QY 350 ---NEFKRD---GRLLTPLGLFDHVTLDKMLAWEPFPIIPRIIRVKDAEAEVAIAKNSD 402
 Db 370 CGGEALKGDFDNGAWVAPVTVTDCSDEMTIVREEIFGFPWMSILSYADEAEVIRANATE 429

RESULT 19
 US-09-252-991A-17138
 ; Sequence 17138, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 17138
 ; LENGTH: 510
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-17138

 Query Match 26.1%; Score 625.5; DB 4; Length 510;
 Best Local Similarity 30.8%; Pred. No. 4.3e-50;
 Matches 152; Conservative 100; Mismatches 205; Indels 37; Gaps 9;

 QY 3 KEYQNVNGEW--KSSVNGIILPIDDSLSGFPAMTREVDFHAKMAGREALPAMAAIT 60
 DB 25 BEOKLYIGRYVASSGATFETINPANGVELAKVQASRESDVERAVQSAGEVKVWAAIT 84
 QY 61 VVERAQYUHKAAIDIERKEELATVLAKEISK--AYNASVTEVVRTADLIRYAAEEGIRL 118
 DB 85 AMQRSKILLRAVDILRRNDELAALETLDTGKPLAETRSV-DIVTGADVLEIYA----- 137
 QY 119 STSADGEGKMDASTGHKLAV-----IRRPQGVILAIPYNYPNVLSGSKIAPALIGG 171
 DB 138 -----GLVPAIEGQIPIRETSEVTRREPLGVVAGIGAWNYPVQIALMKSAPALAA 190
 QY 172 NVVMFKPTQGSGLVLAKAPAEAGLPAGVFNNTITGRSGISGDIYIVHEEVNINTGS 231
 DB 191 NAMIFKPEVTPLTALKAEIVTEAGVPGDFVNLGTSGREVGQWLTEPHLEKISFTGG 250
 QY 232 TPVGQRI---KGLAKWRPIMLEGGKQAGIVLADADLONAAKQIVAGAYDSQRCIAIK 288
 DB 251 TSTGKKWASASSSSUKVETMELGKSFLLIPDPDLDRADIAWMANFFSGGQVCTNGT 310
 QY 289 RVLVVEEVADELAKISENVAKLSVGGPPFD-NATVTPVIDNSADFTSLVVDARQKAK 347
 DB 311 RVFIHRSQQARFEAKVLVQRIQLGDPQDENTNFGVLVSFPHHMESVLGYIESGKAQAK 370
 QY 348 ELMEFKR-----DGLTLTPGLFPHVTLDMKLANEPPGILPILRVKDAEEVALANK 400
 DB 371 LLGCGSERVTDGAFNGAYVAPTFTDCSDMTVREEIFPGVMSILVYDDEBARAND 430
 QY 401 SDGLGSSVFTFRDFOKAFIDANKLEVGVTHNNKTRGFDNFPFLGLKSGAGVGQIRYS 460
 DB 431 TEVGLAGVTVQDLARAHRAHLEAGLCWINT-WGESPAEMPVGKYQSGVGRENGLTT 489
 QY 461 IEAMTNVKSIVLDM 474
 DB 490 LAHYTRIKSVQVEL 503

 RESULT 20
 US-09-134-001C-4541
 ; Sequence 4541, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C

88 AQQSKRMGRFIRBSUGIDUFIMIDIELTA


```

RESULT 27
US-09-733-383-4
; Sequence 4, Application US/09733383
; Patent No. 6664088
; GENERAL INFORMATION:
; APPLICANT: Arjan
; APPLICANT: Narbad, Michael J.C.
; APPLICANT: Rhodes, Michael J.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/733,383
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 09/155,183
; PRIOR FILING DATE: 1997-03-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-733-383-4

Query Match      25.7%; Score 616; DB 4; Length 482;
Best Local Similarity 34.5%; Pred. No. 3.le-49;
Matches 157; Conservative 85; Mismatches 201; Indels 12; Gaps 5

QY      24 SPIDSSLGFVPMATREEVDHAMKAGREALPAAALTVVERAQVLHKAADIIRDKKEIA 83
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      25 NPVTGVELVSRVAATELDADAFAVAQAQFPAAWALLAPNERSLLKAAEQLOARSGEFI 84
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      84 TWLAKEISKAYNASVTEVVRTADLIRYAABEGIRLSTSADEGGKKDASTCKHLAVIRQP 143

```


Db 85 EA-AGETGAMNATYGFVNRVLAANMLREAAS-----MTTQVNGEVIPSDVGFSPAMALROP 138
QY 144 VGVLAIAIPYVPYVNLSSGIAPALGNNVVFVFPKPTQSGVSLVLAFAEAGLPAQVF 203
Db 139 CGVUGIAPWAPVILATRALAMPACNTVVLKASELSPAVHRLIGQVLODAGLGDGVV 198
QY 204 NTITGR--GSEIGDYIVEHEEVNFTGTPVGQIGKLAG--MRPIMLELGGKDGAGI 258
Db 199 NVISNAPADAQIIVERLIANPAVRVNTGTHVGRIVGELSARHLKPALELGGKAPLL 258
QY 259 VLADADLONAKQIVAGDYSGQCTAIKRVLVVEEVADELAEKISENVAKLSVGDPPD 318
Db 259 VLDDADLEAAVQAAGAYFNOCQICMSTERLIVDAKADAFVQAQAKVETLRAGDPAD 318
QY 319 NATVT--PVIDNSADFTESLAVDARQKAGKELNFRDRLITPGLFDHVTLODKLAWEE 377
Db 319 PESVGLSDVASATRIKALIDDAVAKGARLVIGQGLEGSILQPTLDDGVDAEMRLYREE 378
QY 378 PFGPILPIRVKDAEEAVATANKSDFLQSSVFTTRDFQKAPDIANKLEVGTVHNNKTR 437
Db 379 SFGEVAVLVRGEGEALLQLANDSEFGLSAAIFSRDTGRALALAQRVESGICHINGPTVH 438
QY 438 GPNPFFGLKSGAGVQGIKRVISIAMTNVKSIVL 472
Db 439 DEAQMPFGVKSSGSGYFGGKASIEHFTQLRWVTL 473

RESULT 28

US-08-513-841-2

; Sequence 2, Application US/08513841

; Patent No. 5753481

; GENERAL INFORMATION:

; APPLICANT: Niwa, Mineo

; APPLICANT: Saito, Yoshimasa

; APPLICANT: Ishii, Yoshinori

; APPLICANT: Yoshida, Masaru

; APPLICANT: Suzuki, Hiromi

; TITLE OF INVENTION: No. 5753481el L-sorbose Dehydrogenase and No. 5753481el L-sorb

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.

; STREET: 1755 Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MS-DOS Editor

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/513,841

; FILING DATE: 01-NOV-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: UK 9304700.9

; FILING DATE: 08-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 241851/1993

; FILING DATE: 28-SEP-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: NORMAN F. OBLON

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 18-909-0 PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; TELEX: 248855 OPAT UR

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
; US-08-513-841-2

Query Match

25.4%; Score 609; DB 1; Length 497;

Best Local Similarity 30.8%; Pred. No. 1.5e-48;

Matches 149; Conservative 101; Mismatches 213; Indels 20; Gaps 8;

QY 3 KEYQNYVNGEWSKSNQIEILSPIDSSLGFPVPMATREYVDHAKAGREALP--AWAALT 60

Db 14 REFQFFIDGEMRAGKDFDRSSPAHDVPVTRIPRCTREDLDEAVAAARRAFENGWAGLA 73

QY 61 VYEAQVYLKAAADIIRDKBEIATVLAKESIKAYNASVTVETADLIRYAEESGIELST 120

Db 74 AADRAAVLLKAAGLLRERODDIAYWEVLENGKPIQAKGEIDHICFACFEMAGGAARMHLH- 132

QY 121 SADEGGKMDASTGHKL-AVIRROPVGIVLAIAPYNPVNLSSGIAPALIGNNVVMFKPP 179

Db 133 -----GDTFNNLGEGLFGVLRPEFVGVLITPWNFFMILCERAPFLLASGCTLVVKPA 187

QY 180 TQGSVGLVIAKAFAGLPAQVNTITGRSEITGDYIVHEEVNFINFTGTPVQGRIG 239

Db 188 EVTSATTLTLLAILADAGLPGKGVNFVVTGTRTVGQAMTEHQIDMLSTGTGKSCII 247

QY 240 KLA---GMRPIMLEGGKDGAGIVLADADLDNAAQIVAGADYSGORCTAIKRVLVVEEV 296

Db 248 HAAADSNLKKLGLGGLGKNPVIIVFADSNLEORADAVAFISFNTGCCVSSSRILIVERSV 307

QY 297 ADELAKEKISENVAKLSVGDPPDNAT-VTPVIDDNSADPFIESLVVDAROKAKEL-----N 350

Db 308 AEKFERLVVPMKEIRVGDPPDETQIGAITTEAQNKTILDYIAGKAEAGKLLCGGGIV 367

QY 351 EFRKDRLLTPGLFDHVTLODKLAWEEPFGPILPIRVKDAEVAIAANKSDFLQSSVF 410

Db 368 DFGK-GQYQPTFTDVKPSNGIARDEIFGPVLASFHFDVDEAIAIANDTVYGLAASVM 426

QY 411 TRDQKAFDIANKLEVGTVHNNKTGRGPDNPFPLGLKSGAGVQGIKRVISIAMTNVKSII 470

Db 427 SKOLDKALAVTRVRVAGRFVWNTINGSGPET-PLGGFKQSGWRGAGLYGVEETQKSV 485

QY 471 VLD 473

Db 486 HIE 488

RESULT 29

US-08-696-834-2

; Sequence 2, Application US/08696834

; Patent No. 5834263

; GENERAL INFORMATION:

; APPLICANT: Niwa, Mineo

; APPLICANT: Saito, Yoshimasa

; APPLICANT: Ishii, Yoshinori

; APPLICANT: Yoshida, Masaru

; APPLICANT: Hayashi, Hiromi

; TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonic Acid

; NUMBER OF SEQUENCES: 48

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Oblon, Spivak, McClelland, Maier & Neustadt,

; ADDRESSEE: P.C.

; STREET: 1755 Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

```

; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,834
; FILING DATE: 24-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 28612/1994
; FILING DATE: 25-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
; US-08-696-834-2

Query Match 25.4%; Score 609; DB 2; Length 497;
Best Local Similarity 30.8%; Pred. No. 1.5e-48;
Matches 149; Conservative 101; Mismatches 213; Indels 20; Gaps 8;

Qy 3 KEYQYVNGWKSNNQTEILSPIDSSLGFPVPMATREEDVHAKAGREALP--AWAALT 60
Db 14 REFQFFIDGWRAGKDFDRSPAHDPVTRIPRCTREDLDEAVAAARRAFENGSWAGLA 73

Qy 61 VYERAOYLKADIERDKERIAVLAKETSKAVNASVTEVTRTADLIRYAAEGIRLST 120
Db 74 AADRAAVLKAAGLRRERDDIAYWEVLNGXPISQAKGEIDHCIACFEMAGARMHL- 132

Qy 121 SADEGGKMDASTGKHL-AVIRQPVGIVLAIAPYVFNLSGGKIAPALIGNVVMKPP 179
Db 133 -----GDTFNNLCEGLFGWLRPIGVGLITPWNFFPMLICERAPFILASGCTLVVKA 187

Qy 180 TQGSVSGVLAKAFAGLPAQVNTITGRSGEIGDYGIVEHERVFNFTGTSTVPGQIG 239
Db 186 EVTSATLLLAELLADAGLPGVNVVGTGTGQAMTEHQDIDMLSTGTSGVKSCT 247

Qy 240 KLA---GMRPIMLELGGKDAIVLADADLDAKQIVAGAYDYSQRCVTAIKRVLVVEV 296
Db 248 HAADSNLKLGLGLGKNPIVFAADSNLEDAADAVAGISFNTGCCVSSRLIVERSV 307

Qy 297 ADELAEKISNNVAKSVSGDPDNAT-VTPVIDNSADFIESLVVDARQKAKEL-----N 350
Db 308 AEKFERLVWPKMKIRVGDPPDPTQIGAITTEAQNKTILDYAKGAKGAKMLCGGGIV 367

Qy 351 EFKSDGRLTPGLFQHVTLDMKLAWPEPGPILPIIRVKDAEAVAIANKSDFGQSSVF 410
Db 368 DFGK-GYQITLFTDVKPSGIARDEIFGPVLASFHEDTVDEAIAANDTVGLASVW 426

Qy 411 TRDFQKAFDIANKLEVGTVHNNKTGRGPDNFPFLLGLKSGAGVQGIYSIEAMTVKSI 470
Db 427 SKDIDKALATVTRVAGRFWNTIMSGPET-PLGGFKQSGWREAGLYGVETQLKSV 485

Qy 471 VLD 473
```

```

Db 486 HIE 488
;
; RESULT 30
; US-08-942-673-2
; Sequence 2, Application US/08942673
; Patent No. 5861292
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshinasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 5861292el L-sorbose Dehydrogenase and No. 5861292el
; TITLE OF INVENTION: L-sorbose Dehydrogenase Obtained from Gluconobacter
; TITLE OF INVENTION: oxydans T-100
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Obion, Spivak, McClelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,673
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/513,841
; FILING DATE: 01-NOV-1995
; APPLICATION NUMBER: UK 9304700.9
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 241851/1993
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-909-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
; US-08-942-673-2

Query Match 25.4%; Score 609; DB 2; Length 497;
Best Local Similarity 30.8%; Pred. No. 1.5e-48;
Matches 149; Conservative 101; Mismatches 213; Indels 20; Gaps 8;

Qy 3 KEYQYVNGWKSNNQTEILSPIDSSLGFPVPMATREEDVHAKAGREALP--AWAALT 60
Db 14 REFQFFIDGWRAGKDFDRSPAHDPVTRIPRCTREDLDEAVAAARRAFENGSWAGLA 73
```



```

; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4451
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4451

Query Match      25.4%; Score 607.5; DB 4; Length 518;
Best Local Similarity 31.4%; Pred. No. 2.2e-48;
Matches 143; Conservative 100; Mismatches 194; Indels 19; Gaps 8;

QY 9 VNEGWSSVQIEILSPIDSSLT--GFVPAMTRREVDHAKAGREALPAAALTYVERAQY 67
DB 45 INGEKLTGDTFNSFNANTSQIAKYKATQDDIEKAFESANHAYQSKWKSHKORAEI 104
QY 68 LKAAADIIRDEKEEATVLAKEISKAYNASVTEVVTADLIRYAAEEGIRLS TSADEGK 127
DB 105 LLRVAALIRRRKEEISALVYEACKPDEAVGDAEAGIDIEFYARSMMEL---ADGKV 161
QY 128 MDASTGHLAVIRROPVGIIVLAIPYNY--VNLGSKIPALIGNVVMPKPTQGSVSG 186
DB 162 LDREGEHNRVFK--PIGTGVTTPPNPFAIMAGTTLAP--VVGNTVLLKPAEDTVLTA 218
QY 187 LVLAFAFAAGLPAGVFNITGRSGEIGYIVHEEVENFINFTGSPVQRIKGLAG--- 243
DB 219 YKMEILEBAGLPQGVNFWFVPGPKSIGDYLVHDKOHTFVTFTGSRATGTRIYERSAVQ 278
QY 244 -----MRPTMLELGGKAGIVLADADLNAAKQIVAGAYDYGQRCTAIKRLVBEVAD 298
DB 279 EGQOFLKRVIAEWGSKDAIVDNNVDTDLAAEIVTSAGFSQKCSACSRAIVHQDVHD 338
QY 299 ELAEKISENVAKLSVGDFDNATVTPVIDDNSADFISSLVVDARQKAKELNEFKRD--G 356
DB 339 EILEKATQLTQKLTGNTTEENFWGVNQKQFOKIKNYTEIGKKEGKLEGGTGGTDDSTG 398
QY 357 RLLTPGLFDHVITDMKLAWEEPFGPILPIIRVKAEEAIAIAKNSDFGLQSSVTRDFQK 416
DB 399 YFIEPTIFSGLOASDRIMQEEIFGVVGVFKVQFDEAIEVADTDYGLTGAVITNHRH 458
QY 417 AFDIANKLEVGTVHNNK--TCRGDPNFPFLGLKSGG 451
DB 459 WIKAVNEFDVGNLYNRCCTAAVVGHPFGGFKSG 494

RESULT 33
US-09-328-352-5847
; Sequence 5847, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5847
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5847

Query Match      25.3%; Score 604.5; DB 4; Length 498;
Best Local Similarity 33.3%; Pred. No. 3.9e-48;
Matches 160; Conservative 85; Mismatches 215; Indels 21; Gaps 9;

QY 8 YNKGWSSVQ--IEILSPIDSSLTGFVPAMTRREVDHAKAGREA--LPAAALTYVER 64
DB 16 YIAGEWRLGRGAVIQSLFPADQSVNABLSATLIEDVNEAIEKADQAWRQFWRNSLPHER 75

```

```

QY 65 AOYLHKAADIIRDEKEEATVLAKEISKAYNASVTEVVTADLIRYAAEEGIRLS TSADE 124
DB 76 ARILYKADILIEARVDELKSLQTDNGKELTETRLGLVMSAAATARYVA--AACETLNDE 132
QY 125 GGRMDASTGHLAVIRROPVGIIVLAIPYNYPNVNSGSKIAPALIGNVVMPKPTQGSV 184
DB 133 LTTORAPDFWMTMSV--HEPVGVAAITPWNSPIASSEVQKLAPALAGNAVVLKPAEATSL 190
QY 185 SGLVLAFAFAAGLPAGVFNITGRSGEIGYIVHEEVENFINFTGSPVQRIKGLAGM 244
DB 191 TALELAKIFEEAGLPKGLSLVIVGRGSVIGDLAIGHPLVVKISFTGCTTTGSHLAHIAAD 250
QY 245 RPI--MLELGGKAGIVLADADLNAAKQIVAGAYDYGQRCTAIKRLVWVEEVADELA 302
DB 251 KLITTSLELGGKSPVILPVDADVELAAKGVAYGIFSSAGQACIAGSRFIIHSSLYDQFLT 310
QY 303 KISENVAKLSVGDFDNAT--VTPVIDDNSADFISSLVVDARQKAKEL-----NEEK 353
DB 311 ELVEITKGLRVGHPQTCVHLGFLVNDKHLQSVDRYVOLAKSEGGQVLIGGEALTGGDIA 370
QY 354 RDGRLLTPLGLFDHVITDMKLAWEEPFGPILPIIRVKAEEAIAIAKNSDFGLQSSVTRD 413
DB 371 K-GSYVLPITITGLNNSAQTQCEIEIFGVVVMKYDNEQDLAQANDSCFGLAAGIWTES 429
QY 414 FOKAFDIANKLEVGTVHNNKTCRGDPNFPFLGLKSGAGVQGIIRYSIEAMTNVKSIVLD 473
DB 430 YEKAWIRARALEVGTVWINTYK--KFSISAPFGGKDSGIGREKGRGLILSYMOOKSIYM 488
QY 474 M 474
DB 489 L 489

RESULT 34
US-09-976-594-287
; Sequence 287, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; APPLICANT: Furness, Michael
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 287
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2558815CD1
US-09-976-594-287

Query Match      25.2%; Score 602; DB 4; Length 518;
Best Local Similarity 31.1%; Pred. No. 7.2e-48;
Matches 148; Conservative 99; Mismatches 197; Indels 32; Gaps 8;

QY 23 LSPIDSSSL--GFVPAMTR-----EVDHAKAGREALPAAALTYVERAQYLUHK 70
DB 44 VEPADASGTEKAFEPATGRVIATFTCSGEKEVNLAQNAKAAFKINSQKSGMERCILLE 103
QY 71 ADIIRDEKEEATVLAKEISKAYNASVTEVVTADLIRYAAEEGIRLS TSADEGSKMDA 130
DB 104 AARIIREREDEIATMECINNGKSIPEARLDIDISMQCLEYYA-----GLAASMAHEHIQ 157
QY 131 STCHKLAVIRROPVGIIVLAIPYNYPNVNSGSKIAPALIGNVVMPKPTQGSVSLVLA 190
DB 158 LFGSGFGVTRREPLRGVGVIGVGNVYPTQIASWSAFALACGNAMVFKESPFTFVSALLA 217

```

Query Match.	25.1%	Score 601;	DB 4;	Length 506;
Best Local Similarity	32.5%;	Pred. No. 8.6e-48;		
Matches 152;	Conservative 95;	Mismatches 189;	Indels 32;	Gaps 10;
QY	6 QNYVNGEW--KSSVNOIEITLSPIDSSLSGVPWPTREBVDHMKACGREALP--AWAALTV 61	:	:	:
Db	20 QYIDGWEVSSNNTRDILINPYQEIIFVTAEGTKEVDERAILAARSFDEGQWLSYTS 79	:	:	:
QY	62 YERAQYLKAAADIIRKDEKIATVLAKETSKAYNASVTEVVUTADILIRYAABEGRISLTS 121	:	:	:
Db	80 EVRGKKVRADVADIKENREBELAKLETLDGKTLESYADMDIHNVMFYA--GL--AD 134	:	:	:
QY	122 ADEGGKMDASTGKHLAVIRBPYGVILAIAPYNPYNLSGSKTAPALICGNVWFKPPTQ 181	:	:	:
Db	135 KDGGEIINSPINPAESKVVKEPVGVTQITPNYPLLOASWKIAPALATGCSLYMKESEI 194	:	:	:
QY	182 GSVSGILVAKAFEAAGLPAGVNTITRGSEIGDYIVHEHEVNFINTGSPVQGRICKL 241	:	:	:
Db	195 TPLTITRVFELMEVGFPGKLTINVLGSGSEVDVMSGHEBVDLVSTGGTETGKHIKQ 254	:	:	:
QY	242 AG--MRPIMLEIGKGAGIVLADADLDLNAAKQIVAGYDYGQRCTAIVKRVLVVEEVADE 299	:	:	:
Db	255 AANHVTVDVALELGGKFNPIFDQADFELAYDQALQNGFYHAGQVCSAGSRILVINDIKOK 314	:	:	:
QY	300 LAEKISENVAKLSVGPDPFNAD--VTPVIDDNSADFTLSLVVDARQAGKELNEFKDGRLL 358	:	:	:
Db	315 FEKALIDRVKIKUGNGFQDDTEMGFVISTAHNRDKIEGYMEVAKKQDGTATIGKRPERE 374	:	:	:
QY	359 -LTPGLFDHVTL-----DWKLAWEEPFGPIITIRVKKDAEAAVANKSDFGLQSSVFT 411	:	:	:

; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6118
 ; LENGTH: 490
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-6118

Query Match 25.0%; Score 598; DB 4; Length 490;
 Best Local Similarity 32.1%; Pred. No. 1.6e-47;
 Matches 151; Conservative 95; Mismatches 210; Indels 14; Gaps 6;

QY 12 EWKSSVNOIEILSPIDSSIGFVPAMTREVVDHAKGREALPAAVTVVERAQLVHKA 71
 DB 26 EAPASNDYEVNDAANGTETLAWKTYDRAGVEAIAINSAQAQAAKKKTALARAADVLLAW 85
 QY 72 ADIIRDKKEIATVLAKIEISKAYNASVTEVVRTADLIRYAAEEGIRLSTSADEGKNDAS 131
 DB 86 YNLMLEKENLAQILTAEOCKPLAEARGEIGYAAASFIRWFAEQARRI-----DGEVLTPT 140
 QY 132 TGHKLAVIRQVGVILAIAPYVNVNLSGSKIAPALIGNVVWFKPTQSGVSLVLAK 191
 DB 141 LPNQLLVIKQAIQVTAITPWNFPAAMITRKAGPAIAAGCSMLVKPAEQTLTVALEV 200
 QY 192 AFAEGLPAGVENTITGRSGEIGDYIVEHEEVNFTGTPVQRIKLAG--MRPIML 249
 DB 201 LAQAGLADVLINISGNAVEKTLCSDIVRKLSITGSTQVGRILMQCAPTIKKLSL 260
 QY 250 ELGCKAGIVLADADLDAKQIVAGDYSGQKCTAIRKVLVBEVDELAELAEKISNVA 309
 DB 261 ELGNAFVVFDANLEQAVQGWIMASKYRNSGQTCVCANRIYVQDGIYDALADLREAVS 320
 QY 310 KLSVGD--PFDNATVTPVIDNSADFTESLVVDARQKAKELNFKRD---GRLTLPGLFD 365
 DB 321 KLVGDGRGSGTQGLIDELDAIKVQSHIADATEKGAIVRIGGRKESALGTFPEFTVLT 380
 QY 366 HVTLDKLAWEPPGPIPIIRVKDAEVAIAANKSDFGLOSVFTFTRDFQKAPDIANKLE 425
 DB 381 GVTQDMKVKSETFGPLAPLFRFKTEDEAVAMANDTFGLAAVLFTQSTARQWVGEAL 440
 QY 426 VGTVHNNKTR--GPDNFFPLGLKSGAGVQGIYSIEMTNVKSIVLDM 474
 DB 441 YGMVGIN--TGALSNVAFPGVYKQSLGRSGKFGIEEYVEMKYLVDL 488

RESULT 38

US-09-717-926-2
 ; Sequence 2, Application US/09/717926
 ; Patent No. 6569657
 ; GENERAL INFORMATION:
 ; APPLICANT: Meyers, Rachel
 ; APPLICANT: Cook, William J.
 ; TITLE OF INVENTION: 32140, A No. 6569657el Human Aldehyde
 ; TITLE OF INVENTION: Dehydrogenase
 ; FILE REFERENCE: 35800/205243
 ; CURRENT APPLICATION NUMBER: US/09/717,926
 ; CURRENT FILING DATE: 2000-11-21
 ; PRIOR APPLICATION NUMBER: US 60/214,707
 ; PRIOR FILING DATE: 2000-06-27
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 923
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-717-926-2

Query Match 25.0%; Score 597.5; DB 4; Length 923;
 Best Local Similarity 30.8%; Pred. No. 4.8e-47;
 Matches 152; Conservative 97; Mismatches 207; Indels 37; Gaps 13;

QY 5 YQYVNGEMKSSVN--QTEILSPIDSSIGFVPAMTREVVDHAKGREALP--AWAALT 60
 DB 443 YQCFINGQFTDADGKTYDTINTDGTSTICKVSVASLADVDAKAAKDAFENGWGRMN 502
 QY 61 YVERAQLVHKAADIIRDKKEIATVLAKIEISKAYNASV--TEVVRTADLIRYAAE--EGIR 117
 DB 503 ARERGRIMYRLADLLEENQBELATIEALDSGAVVYTLAKTHIGMSVOTFRYPAGWCCKIQ 562
 QY 118 LSTSADEGGKMDASTGHKLAVIRROQVGVILAIAPYVNVNLSGSKIAPALIGNVVMEK 177
 DB 563 GSTTIPN---QARPNRMFTTKKEPLGVCAIITPWNPLMLLAWASACLAAGNTLVLK 618
 QY 178 PPTQSGVSLVLAKAPAPAGPAGVFNITITGRSGEIGDYIVEHEEVNFTGTPVQGR 237
 DB 619 PAQVTPJTALCKPAELSVKAGFPKGVNIIIPGSGGIAGORLSEHPDIRKLGFTGTPIGKQ 678
 QY 238 TCK---LAGMRPIMLELGCKDAGIVLADADLDAKQIVAGDYSGQKCTAIRKVLV 294
 DB 679 IMKCAVSNIKKVSLELGKSPLIIFNDCELDKAVRMGMGAVFNNKGENCIAAGRLFVEE 738
 QY 295 EYADELAEKISENVAKLSVGDPFDNATVTPVID---DNSADFTESLVVDARQKAKE--- 348
 DB 739 SHDFEFTVRVVEIKKMKIGDPLDRST-----DHGPNHKAHLEKL--LOYCETGVKEGAT 792
 QY 349 ----LNEFKEDGKSLTTPGLFDHVTLDKLAWEPPGPIPIIRVK--DAEBAVAIAKSD 402
 DB 793 LVYGGQVQRPQGFMEPTVFTDVEDYMYLAKEESTFGPTMVISKFQNGDIDGVLOKANSTE 852
 QY 403 FGLQSVFTEDQKAFDANKLEVGVTHLN--NKTGRGPDNFFPLGLKSGAGVQGIYS 460
 DB 853 YGLASGVFTDINKAMYSEKLEAGIVFINTNKTDVAA---PFGVKSGGFKDLGEEA 909
 QY 461 TEAMTNVKSIVLD 473
 DB 910 LNEYLTKTITL 922

RESULT 39

US-09-328-352-7337
 ; Sequence 7337, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 7337
 ; LENGTH: 505
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-7337

Query Match 24.8%; Score 594; DB 4; Length 505;
 Best Local Similarity 31.1%; Pred. No. 3.9e-47;
 Matches 153; Conservative 109; Mismatches 198; Indels 32; Gaps 11;

QY 1 LTKYQNVNNGEKSS--VNQTEILSPIDSSIGFVPAMTREVVDHAKGREALP-AWA 57
 DB 24 MARQHGHPIDGKLVAALDLKVDIVNPESTEQVVAQISGQQDVESAVKSAHAFQAWA 83
 QY 58 ALTVYERAQLVHKAADIIRDKKEIATVLAKIEISKAYNAS--VTEVVRTADLIRYAAEEGI 116
 DB 84 EFTPYERGKVLKADLADLEQYGEELAOLETSTGLKINISHLEVAQSVIFLRYFAGWAT 143
 QY 117 RL-----STSADEGGKMDASTGHKLAVIRROQVGVILAIAPYVNVNLSGSKIAPALI 169
 DB 144 KINGQTWPQSPISMQGEXKYTFT-----LRQPVGVAGVFPWNFSLMIGVWKIGSALT 196
 QY 170 GNVVWFKPTQSGVSLVLAKAPAEAGLPAGVFNITITGRSGEIGDYIVEHEEVNFT 229

Db 197 TCGTIVLKPGEFASLSILRLAELAIEAGIPAGVINVTGKG-DTGQVLLSPVKKVSFT 255
 QY 230 GSTPVGQRIGKLA---GMPIMLELGGKAGIVLADADLNAKQIVAGAYDSQORTA 286
 Db 256 GSVPTGAIGKLA MSDLTRVSLGELGGKNAIADANIDILPILLOATFVHQGVAS 315
 QY 287 IKRVLVVEVADELAEKISENVAKLSVGDPDFDNATV-TPVIDDSDAFIESLVDAKQK- 344
 Db 316 PERFVHTKYDELVDKLSKALSKQKSGSAMDEGSMFGLSNQPHFKVXHYLDMAKANN 375
 QY 345 ----GAKELNFRKDRGLLTPGLFDHVTDMKLAEEPPGPIPIIRVKDAEAEVAIAK 400
 Db 376 QIIAGGALDQI---GVFQPTLISFKNTDDPLFSSETFGPVGVMFPFDEBELIQLANQ 432
 QY 401 SPGLQSSVFRDQKAFIANKLEVTGVHNNKTGRGPDNFPPLGLKSGAGVQGIYS 460
 Db 433 SRFGLTASITWINDLSKALRLIKIEAGTLWVNMHTLDP-SVPFGGVKASGIGRFSDAF 491
 QY 461 IEAMTNVKSIVL 472
 Db 492 IEDYELKSVMI 503

RESULT 40

US-09-252-991A-28449
 ; Sequence 28449, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 28449
 ; LENGTH: 748
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-28449

Query Match 24.7%; Score 591.5; DB 4; Length 748;
 Best Local Similarity 30.3%; Pred. No. 1.3e-46;
 Matches 148; Conservative 91; Mismatches 215; Indels 35; Gaps 7;
 QY 6 QNYVNGEWSVNOQIEILSPIDSSIGFVPAMTREEDVDMKAGREALPAAWALTIVVERA 65
 Db 265 KHMINGREVESKOVFNPNPATGELIGEVAGSGAAEIDAANAAREAPFKWANTPAKERA 324
 QY 66 QYHLKKAADIIRDEKEELATVLAKEISKAYNASVTEVV-RTADLIRYAAEEGIRLSADE 124
 Db 325 RLMRLEGELIDRNVPHLAELTDLTGLPIHQTKNVLPRAHNFEEFAEVCTRWN----- 379
 QY 125 GKXMDASTGHKLAVIR-----QVGVILAIAPYNPVNLGSKIAPALIGNVNMF 177
 Db 380 -----GHSYFVDDQMLNLTLYQVGVCGVLSPFMVNPMTATWKTAPCALGNATVILK 431
 QY 178 PPTQGSVGLVLAFAFAEAGIPAGVENTITGRGSEIGDYIVEHEEVNFINTGTPVGOR 237
 Db 432 MSELSPUTANEGLRVLHEANGIPGVNVVQGVASAGDALVRHDRVAVSTGTATGR 491
 QY 238 IGKLAGMRPIMLELGGKAGIVLADADLNAKQIVAGAYDSQORTAIXRVILWVEVA 297
 Db 492 IWEAAGIKKYSMELGKSPVLVFDADLERALDAALFTIFSINGERCTAGSRIFVQESVY 551
 QY 298 DELAEKISENVAKLSVGDPDFDNAT-----VTPVIDDSDAFIE-----SLVDAKQK 345
 Db 552 PQFVAEPAARARLLVGDPPDKTQVSGMITQAHYDKYVTGYIRIGLEEGATLVAGGLERP 611

QY 346 AKELNEFRKDRGLLTPGLFDHVTDMKLAEEPPGPIPIIRVKDAEAEVAIAKSGDFGL 405
 Db 612 A-GLPAHLKSKQFTQTPVADVDNRMRIAQSEIFGPPVCLIPFKDAEALRLANDVEYGL 670
 QY 406 QSSVFRDQKAFIANKLEVTGVHNNKTGRGPDNFPPLGLKSGAGVQGIYSIEAMT 465
 Db 671 ASYIMTQDIGHKARLARGIEAGVTFINSQNVRLRQ-PFGGVKASGTGRGGEYSFEVA 729
 QY 466 NVKSIIVLDM 474
 Db 730 EIKNVICSM 738

RESULT 41

US-09-328-352-6757
 ; Sequence 6757, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6757
 ; LENGTH: 500
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-6757

Query Match 24.7%; Score 590.5; DB 4; Length 500;
 Best Local Similarity 33.2%; Pred. No. 8.1e-47;
 Matches 161; Conservative 76; Mismatches 225; Indels 23; Gaps 10;
 QY 1 LTRKEYQNYVNGEWSVNOQIEILSPIDSSIGFVPAMTREEDVDMKAGREAL--PAAW 58
 Db 12 MTADFOLYINGQFESGAATFESINPATGEWAQMFPEARTDEVNRAVQASOALKASAWAG 71
 QY 59 LTVYERAOYLHKAADIIRDEKEELATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRL 118
 Db 72 LTASQGRKLYKLADLIEKAAPQIAQIETSDGTGLIETSSQIAYVAEYRYVA--GI-- 127
 QY 119 STSAD--EGKMDASTGHKLAVIRQFVGIVLAIAPYNPVNLGSKIAPALIGNVNMF 176
 Db 128 ---ADKLEGSFLPIDKADQAWIVREPVGVAALIVPWSOLFSAVKVGPALAAAGCTVWL 184
 QY 177 KPPTQGSVGLVLAFAFAEAGIPAGVENTITGRGSEIGDYIVEHEEVNFINTGTPVGQ 236
 Db 185 KASBDGFGPILAFALVHEAGFPAGVNVNITGFGECGAVLSHPDVAHVAPGTGPETAR 244
 QY 237 RI--GKLAGMRPIMLELGGKAGIVLADADLNAKQIVAGAYDSQORTAIXRVILWVE 294
 Db 245 HIVNSAENLAKVLSLELGGKSPFIVPADADIQSAVNAQVAIAFAATGSCVAGSRLVVEE 304
 QY 295 EYADELAEKISENVAKLSVGDPDFDNAT-VTPVIDDSDAFIESLVDAKQKAGEL---N 350
 Db 305 SIKDEFQVRIVERVQVTIKIGLPHEMATEFGPLCTLRQKQIEQVWASSIQOAGKALLTGGK 364
 QY 351 EFKDGRGLLTPGLFDHVTDMKLA--WPEPGPIPIIRVKDAEAEVAIAKSGDFGLQS 407
 Db 365 SLERAGYYPTTILD--CTDVPQADCTITELFGPVLSDVDFKDEAEAVOKANSTPYGLAA 422
 QY 408 SVFTFRDQKAFDIANKLEVTGVHNNKTGRGPDNFPPLGLKSGAGVQGIYSIEAMTNV 467
 Db 423 GVFTTNLSRAHRTKVIGSVGLVNLTYRAVSP-LAPFGHGLSHGREGGANAVLDYTTT 481
 QY 468 KSIIVL 472
 Db 482 KIVML 486

RESULT 42

Qy	3	KYQYVNVNGEKSVNQVIELLSPIDDSGLGTFVPMTRBREVVDHAMKAGRALPAAWALTUY	62
Db	6	KKINHVKNKVAGADYFHTTNPATGVEIAEVAASGEAEINQAAAKEAFPKWANLPMK	65
Qy	63	ERAQYLHKAADILIERKEBIAVLAKESIKAYNASVTEW-RTADLIRYAAEBEGIRLUSTS	121
Db	66	ERARLWRRLGDLDDQNVPEIAMETADTGLPHQTKNVLPRASHNFEFAE-----VC	119
Qy	122	ABEGKMDASTGHKLAIVIRROPVGIUVAIAPNYPVNIUGSKTAPALIGNVNVFKPPTQ	181
Db	120	QQMNGKTYFVDKMLNYTLVQPVGCALVSGPMWNPMTATWKVAPCALGNTATVLKNSL	179
Qy	182	GSVSGVLAKAFEAEGALPAGVFNITGRGSETGDVIVHEBVNFITGSTSPVCGRTGL	241
Db	180	SPLTADRRLGELALEAGIPAGVLNVVQVGGATAGDALVEHHDVRAVSGTGTATGNIMKN	239
Qy	242	AGHRPIMIELGGKDGAGIUAADADLNAAKOIVAGAYVSGQRTATKRVLVVEEVADELA	301
Db	240	AGLKYSMELGGKSPVLIFEDADIERALDAALFTFSINGERCTAGSRITFQOSIYPEFV	299
Qy	302	EKISENVAKLSGDPDF-NATVTPVIDNSADFIESLVUDAROKGA-----KELN	350
Db	300	KRFAPRANRLVGDPIDPTQVGLISQOHWKVSGYTRLGEIEGATLLAGGAKPSPDLP	359
Qy	351	EFKRPDRLLTFLGFDHVTLDMLKLAWEBSFPFILPIIRYKDABEAVAANKSDFCLOSSFV	410
Db	360	AHLKAGNFRPTVLAVDVNRMEVAQESIFGPVACLLPKFQAEGLRLANDVEYGLASYIW	419
Qy	411	TRDFOKAFDIANKLEVGTVHNNKTGSDPNPFELGLKSGAGVQOIGRYSIEAMTNVKSI	470
Db	420	TQDVSKVRLRARGIEAGMVFVNTQVNRDLBQ-PFGVQASGTGREGGEYSFEVFAEMKN	478
Qy	471	VLDM	474
Db	479	CISM	482

```

RESULT 45
US-09-489-039A-13971
; Sequence 13971, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13971
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13971

```

```

Qy 166 PALIGNVNMKPPQTQSGVSLVLAKAFABAGLIPAGVNTIITGRSGEIGDVIHEBEVNF 225
Db 189 PALAACSGTVPILKPSETPTLLTLEVAELATQAGIPDGVFNVVTGSGAGCAALTAHPQVAK 248
Qy 226 INFTGSTPVGQRIKGLAGMR--PIMLELGGDKAGIVLADADLDNAAKQIVAGAYDYGQR 283
Db 249 VSTGSTATGKQIARVADRLTFTVLELGGKNPAIVLKQADPQWVIEGLMTGSFLNQGV 308
Qy 284 CTAIKVVLVVEBEVADELAEKISNNAKUSVGDPF-DNAVTVPTVDNDSADFIESLVVDAR 342
Db 309 CAASSRIYTEAPLEFTLVSGFEQAVKSLQVPGMQETAQINPVVSRAHCDKVAAYLEEAR 368
Qy 343 QKGAKELENEF--KEDRELITPGLDPHVITLDMKLAWEERPGPIPIIRVKDABEVAIAIN 399
Db 369 QOKRAELISSGADPGAGGYIPTLVNPDAGRLSREVEFGPVNVLVRVADGEELRLAN 428
Qy 400 KSDFGHQSSVTRDFQKAFDIANKLEVGTVHNNKTRGPDNPPFLIGLKSGGAG 453
Db 429 DSDFGLTASVTRDLITQALNYTDRLQAGTVVNSHT-LIDANLPFGMKQSGTG 481

RESULT 46
US-09-328-352-4475
; Sequence 4475, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4475
; LENGTH: 481
; TYPE: PET
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4475

```

```

US-09-489-039A-13971
; Sequence 13971, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/03/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13971
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13971

```


ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25398

Query Match 23.48; Score 561; DB 4; Length 642;

Best Local Similarity 31.74; Pred. No. 7.2e-44;

Matches 150; Conservative 79; Mismatches 186; Indels 58; Gaps 12;

QY 31 LGFVPAMTRE---EVDHAMKAGREALPAAALTVYVERAQYLHKAADIIERDKEETATVL 86
DB 166 LGSTPLLDADAAALAAALDAARAYDHGGGLWENLEVERIQHVERELARMEQREAVVKLL 225
QY 87 AKBIKXANNASVTEVVTADLI-----RYAABEGIRLSTSADEGGKMDAS 131
DB 226 MWEIGKNLKDSEKEFDRCTDYIVDTIETALKELDRSSRFELRQ----- 269
QY 132 TGHKLAVIRROPVGVILATAPNYPNLSGSKTAPALIGNVVMKPKPTOGSVSGVLAK 191
DB 270 ---TLQIRRPVLGVALCMGFYNYPLNETFTLLIPALIMGNTVWFPAKXGVILLIRPLLE 326
QY 192 AFAEAGLPAGVFNITIGRSGEIGDYIVEHEEVNFTGSTEIVGQRIKLAGMRP---I 247
DB 327 APRDS-FPGVINVIYGRRETVSALMASGVDFVAFIETHSGASDLKLL-HRPRLRA 384
QY 248 MLELGKAGIVLADADLNAAKQIVAGDYSGQRTAKRVLVVEEVADELAEKISEN 307
DB 385 ALGLDAKNPGLVQVLDLNAVNEAVTGALSFNGQRTAKILFVHESVLRPLDKFSAR 444
QY 308 VAKLSVGDFFD-NATVTPVIDDNSADFTIESLVVDARQKAKELNEFKEDGR--LLTPGLF 364
DB 445 LASLKPQMFWEFGVALTPPEQKVDYLEGLLNDALAKARVYNAGGGEHRQSFYPALL 504
QY 365 DHVTDMKLAMEEPFGPILPIRVKDAEAAVAITANKSDFLQSSVFTRD-----FQKAF 418
DB 505 SPVSPDMRLYHEEQGELVVPVYRELDVIVLSRQQLSLFGNDPAQIGRLVDAP 564
QY 419 DIANKLEVGTTHINKTGRGPNFPFLGKSGAGVQGIYRIEAMTNVKSIV 471
DB 565 --AN--QVGRINVAOCQPGDPSYPFNGRKNABGTMSVHDAIRVF-SIRTLV 612

RESULT 50

US-09-328-352-7905

; Sequence 7905, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 7905

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-7905

Query Match

Best Local Similarity 30.84; Score 555; DB 4; Length 564;

Matches 151; Conservative 93; Mismatches 208; Indels 38; Gaps 13;

QY 9 VNGE---W-----KSSVNOIEILSP---IDSSLSGLFVPAMTREVVDHAMKAGREAL----P 54
DB 56 VDSELKLWAGATKKTLSPIWQQPDGSLQVQLGSPVWGEKESDEALEAARVYNNRG 115
QY 55 AWAALTVYVERAQYLHKAADIIERDKEETATVLAKESKAYNASVTEVVTADLIRYAAEE 114
DB 116 EWPMMKVSEIACQNFQIRWVEQRDLLIKLIMWEIGSLADSEKEKEFDITITWR---Q 171
QY 115 GIRLSTSADEGGKMDASTGHKLAVIRROPVGVILATAPNYPNLSGSKIAPALIGNVV 174
DB 172 TIDALKDLNANSRFFVIAEQT:GQIRRTPLGLVLCMGPNYPLNBTATLIPAMLMGNTI 231

QY 175 MFKEPTQGSVGLVLAKAFABAGIPAGVENTITGRSGEIGDYIVEHEEVNFTGSTPV 234
DB 232 IFKPDQFGTLLFEPLLEAFRDS-FPKGVNITLIYAPGSLVVPHLASGQINVLIGSKV 290
QY 235 GQRIGKLAGMRP-----IMLELGGKDGAGIVLADADLLENAKQIVAGAYDYSGQRTAIK 288
DB 291 ADHLKK---QHPKSHRLFAILGLDANKNAAILPDAADLTLVKECLLGALSFGQRTALK 347
QY 289 RVLYVEEVADELAEKISENVAKLSVGDFFD-NATVTPVIDDNSADFTIESLVVDARQKAK 347
DB 348 MLMVHRSIADFEVNLRTTELAKLVCMPEKGVSIITPLQGMHRTAYMTEVIEDAVAKGAK 407
QY 348 ELN----EFKRDGRLLTPGLFDHVTLDMLKLAWEPPFGPILPIRVKDAEAAVAITANKSDF 403
DB 408 VVNPGESECK--TMFPAVAVYVTEGMELIREQFGVVPVAVVDDIETVLDVTTSDH 465
QY 404 GLOSSVFTTRDFQKAFDIANKL--EVGTVHINKTGRGPNFPFLGKSGAGVQGIYRI 461
DB 466 GQOVSIFGSDPAQIGHLVDTLVHQVCRVNNINCCQCRGPDVFPFGGRKDSABGTILSVHDA 525
QY 462 EAMTNVKSIV 471
DB 526 RAF-SIRSMI 534

Search completed: April 13, 2004, 14:23:34

Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:08:38 ; Search time 59 Seconds
(without alignments)
2274.747 Million cell updates/sec

Title: US-09-868-195-12
Perfect score: 2393
Sequence: 1 LTKYQNVNGEKKSVNVI.....GIRYSIEATWVKSVILDMK 475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 50 summaries

Database :

A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2382	99.5	475	5	ABP27643 Streptoco
2	2380	99.5	475	3	AAY96743 Streptoco
3	2373	99.2	475	5	ABP29820 Streptoco
4	1812	75.7	474	6	ABU01538 S. pneumo
5	1812	75.7	474	6	ABP81589 Streptoco
6	1802	75.3	474	6	ABU46050 Protein e
7	1700	71.0	475	6	ABU44392 Protein e
8	1699	71.0	439	4	AAU37723 Streptoco
9	1641	68.6	475	5	ABP27644
10	1639	68.5	475	6	ABU46701 Protein e
11	1397	58.4	486	6	ABU25619 Protein e
12	1388.5	58.0	486	6	ABU24815 Protein e
13	1383	57.8	482	6	ABU24149 Protein e
14	1172.5	49.0	498	3	AAG40912 Zea mays
15	1169.5	48.9	496	7	ABM74113 DNA clone
16	1168.5	48.8	496	3	AAG32607 Arabidops
17	1168.5	48.8	496	3	AAG49550 Arabidops
18	1164.5	48.7	496	3	AAG05753 Arabidops
19	1128.5	47.2	461	3	AAG32608 Arabidops
20	1128.5	47.2	461	3	AAG49551 Arabidops
21	1103.5	46.1	438	3	AAG49552 Arabidops
22	1103.5	46.1	438	3	AAG32609 Arabidops
23	1102.5	46.1	438	3	AAG40913 Zea mays
24	1099.5	45.9	438	3	AAG05754 Arabidops
25	1036.5	43.3	402	3	AAG05755 Arabidops

26	868.5	36.3	475	6	ABU48897 Protein e
27	786.5	32.9	492	6	AAE37093 Human 913
28	784.5	32.8	293	3	AAG40914 Zea mays
29	777	32.5	493	7	AAE39889 Human ald
30	741.5	31.0	484	4	AAB79356 Corynebact
31	741.5	31.0	484	4	AAG92676 C glutami
32	726	30.3	482	4	AAU34671 E. coli c
33	726	30.3	482	6	ABU28726 Protein e
34	721.5	30.2	483	6	ABU19008 Protein e
35	721.5	30.2	488	2	AAE55135 Heat resi
36	718	30.0	488	5	ABBA47807 Listeria
37	718	30.0	488	6	ABU32914 Protein e
38	711	29.7	482	6	ABU31620 Protein e
39	705.5	29.5	491	5	ABBO8382 PcuC amin
40	705.5	29.5	491	6	ABU2062 Pseudomon
41	704	29.4	480	6	ABU40285 Protein e
42	693	29.0	480	6	ABU41623 Protein e
43	693	29.0	480	6	ABU42152 Protein e
44	692	28.9	482	6	ABU47293 Protein e
45	690	28.8	489	6	ABU23300 Protein e
46	682	28.5	482	4	AAU38454 Salmonell
47	682	28.5	482	6	ABU47996 Protein e
48	675.5	28.2	493	3	AAG23559 Arabidops
49	675.5	28.2	495	3	AAG23558 Arabidops
50	674.5	28.2	495	3	AAG51350 Arabidops

ALIGNMENTS

RESULT 1
ABP27643
ID ABP27643 standard; protein; 475 AA.

AC ABP27643;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 4462.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

FN WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

XX 24-NOV-2000; 2000GB-00028727.

XX 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masighani V, Margarit Y Rosi, Grandi G, Fraser C;

XX Tettelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABN68274.

XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.

XX Claim 1; Page 3607; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 SQ Sequence 475 AA;

Query Match 99.5%; Score 2382; DB 5; Length 475;
 Best Local Similarity 99.2%; Pred. No. 4.7e-190;
 Matches 471; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LTKYQNTVNGEKKSSVNOIETLSPIDDSGLGFPVPMATREVDHAKGREALPAWAALT 60
 Db :|||||
 1 MTKYQNTVNGEKKSSVNOIETLSPIDDSGLGFPVPMATREVDHAKGREALPAWAALT 60
 QY 61 VYERQYLHKAADI IERDKEETATVLAKEISKAYNASVTEVTRADLIYAAEGIRLST 120
 Db :|||||
 61 VYERQYLHKAADI IERDKEETATVLAKEISKAYNASVTEVTRADLIYAAEGIRLST 120
 QY 121 SADEGGKMDASTGHKLAVIRQPVGIVLAIPYVNLGSKTAPALIGGNVWFKEPPT 180
 Db :|||||
 121 SADEGGKMDASTGHKLAVIRQPVGIVLAIPYVNLGSKTAPALIGGNVWFKEPPT 180
 QY 181 QSVSGSLVLAKAFABAGLPAAGVNTITGRSGEIGDYIVVEHEEVNFTGSPVQRIK 240
 Db :|||||
 181 QSVSGSLVLAKAFABAGLPAAGVNTITGRSGEIGDYIVVEHEEVNFTGSPVQRIK 240
 QY 241 LAGMRPIMLEIGGKDGAGIVLADLDNAAKQIVAGAYDSQRCCTAIKRVLVVEVADEL 300
 Db :|||||
 241 LAGMRPIMLEIGGKDGAGIVLADLDNAAKQIVAGAYDSQRCCTAIKRVLVVEVADEL 300
 QY 301 AEKISENVAKLSVGDPPFNATVTPVIDDNSADFIISLVVDARQKAGKELNEFKRGRLLT 360
 Db :|||||
 301 AEKISENVAKLSVGDPPFNATVTPVIDDNSADFIISLVVDARQKAGKELNEFKRGRLLT 360
 QY 361 PGLFDHVTLDKLAWEPPPGTLPILIRVKDAEVAIAKSDFGLOSSVFTTRDFQAFDI 420
 Db :|||||
 361 PGLFDHVTLDKLAWEPPPGTLPILIRVKDAEVAIAKSDFGLOSSVFTTRDFQAFDI 420
 QY 421 ANKLEVGTVHNNKTGRGPNFPFLGLKSGAGVQGIYSIEAMTNVKSIVLDMK 475
 Db :|||||
 421 ANKLEVGTVHNNKTGRGPNFPFLGLKSGAGVQGIYSIEAMTNVKSIVLDMK 475

RESULT 2

AA96743
 ID AA96743 standard; protein; 475 AA.

AC AA96743;
 XX
 DT 26-SEP-2000 (first entry)
 XX Streptococcus NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
 DE
 XX Group B Streptococcus; MS10; non-phosphorylating; NADP-dependent;
 KW glyceraldehyde-3-phosphate dehydrogenase; NPGAP-3-DH; virulence;
 KW antibiotic; vaccine; gene therapy; veterinary; mastitis.
 XX
 OS Streptococcus agalactiae.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 297

FT
 FT Misc-difference 439 /note= "Encoded by GCW"
 FT
 XX /note= "Encoded by CCW"
 PN W0200037490-A2.
 XX
 XX 29-JUN-2000.
 XX
 PF 22-DEC-1999; 99WO-GB004376.
 XX
 PR 22-DEC-1998; 98GB-00028346.
 PR 20-JAN-1999; 99GB-00001233.
 PR 20-JAN-1999; 99GB-00001234.
 PR 12-APR-1999; 99GB-00008321.
 PR 24-MAY-1999; 99GB-00012036.
 PR 23-SEP-1999; 99GB-00022596.
 XX
 XX (MICR-) MICROSCIENCE LTD.
 PA
 XX Hughes WG, Santangelo JD, Lane JD, Feldman R, Moore JC;
 PI Everest P, Dobson RJ, Henwood CU, Dougan G, Wilson RK;
 XX
 DR WPI: 2000-442636/38.
 DR N-PSDB; AAA51293.
 XX
 PT Peptide encoded by operon including genes identified and obtained from
 PT group B Streptococcus, for treating streptococcal infections.
 XX
 PS Example 5; Page 31-32; 32pp; English.
 XX
 CC The S. agalactiae MS14 gene encodes a protein which has homology with a
 CC non-phosphorylating, NADP-dependent glyceraldehyde-3-phosphate
 CC dehydrogenase (NPGAP-3-DH). Peptides derived from group B Streptococcus
 CC (Streptococcus agalactiae) proteins encoded by genes MS4, MS10, MS11,
 CC MS14 and/or MS16 are claimed. The peptides are useful for screening
 CC potential drugs, or for the detection of virulence, and for the
 CC manufacture of a medicament for use in the treatment or prevention of
 CC infections such as focal infection (including osteomyelitis, septic
 CC arthritis, abscesses and endophthalmitis) and urinary tract infections
 CC caused by group B Streptococcus. The peptides and vaccines comprising the
 CC peptides are useful in treatment of chronic mastitis, especially in cows
 CC (i.e. for veterinary purposes)
 XX
 SQ Sequence 475 AA;
 Query Match 99.5%; Score 2380; DB 3; Length 475;
 Best Local Similarity 99.6%; Pred. No. 6.9e-190;
 Matches 473; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 LTKYQNTVNGEKKSSVNOIETLSPIDDSGLGFPVPMATREVDHAKGREALPAWAALT 60
 Db :|||||
 1 LTKYQNTVNGEKKSSVNOIETLSPIDDSGLGFPVPMATREVDHAKGREALPAWAALT 60
 QY 61 VYERQYLHKAADI IERDKEETATVLAKEISKAYNASVTEVTRADLIYAAEGIRLST 120
 Db :|||||
 61 VYERQYLHKAADI IERDKEETATVLAKEISKAYNASVTEVTRADLIYAAEGIRLST 120
 QY 121 SADEGGKMDASTGHKLAVIRQPVGIVLAIPYVNLGSKTAPALIGGNVWFKEPPT 180
 Db :|||||
 121 SADEGGKMDASTGHKLAVIRQPVGIVLAIPYVNLGSKTAPALIGGNVWFKEPPT 180
 QY 181 QSVSGSLVLAKAFABAGLPAAGVNTITGRSGEIGDYIVVEHEEVNFTGSPVQRIK 240
 Db :|||||
 181 QSVSGSLVLAKAFABAGLPAAGVNTITGRSGEIGDYIVVEHEEVNFTGSPVQRIK 240
 QY 241 LAGMRPIMLEIGGKDGAGIVLADLDNAAKQIVAGAYDSQRCCTAIKRVLVVEVADEL 300
 Db :|||||
 241 LAGMRPIMLEIGGKDGAGIVLADLDNAAKQIVAGAYDSQRCCTAIKRVLVVEVADEL 300
 QY 301 AEKISENVAKLSVGDPPFNATVTPVIDDNSADFIISLVVDARQKAGKELNEFKRGRLLT 360
 Db :|||||
 301 AEKISENVAKLSVGDPPFNATVTPVIDDNSADFIISLVVDARQKAGKELNEFKRGRLLT 360

QY 361 PGLFDHVTLDMLKAWEPFGPILPIIRVKDAEEAVALANKSDFGLOSSVTRDFOKAFDI 420
 DB |||||
 361 PGLFDHVTLDMLKAWEPFGPILPIIRVKDAEEAVALANKSDFGLOSSVTRDFOKAFDI 420
 |||||
 QY 421 ANKLEVTGTHINNKTRGPNPFPLGLKSGAGVQGIYSIRSIEMTNVKSIVLDMK 475
 DB |||||
 421 ANKLEVTGTHINNKTRGPNPFPLGLKSGAGVQGIYSIRSIEMTNVKSIVLDMK 475
 |||||

RESULT 3

ABP29820
 ID ABP29820 standard; protein; 475 AA.
 XX
 AC ABP29820;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 8816.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB004789.
 XX
 PR 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;

Tettelin H;

WFI; 2002-352536/38.

DR N-PSDB; ABN70451.

XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX

Claim 1; Page 3999; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins

Sequence 475 AA;

Query Match 99.2%; Score 2373; DB 5; Length 475;

Best Local Similarity 98.9%; Pred. No. 2.7e-189;

Matches 470; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LTKEYQNYNGKSSVNOIEILSPIDSSSLGFPAMTREVEVDHAKAGREALPAWALT 60
 DB |||||
 1 LTKEYQNYNGKSSVNOIEILSPIDSSSLGFPAMTREVEVDHAKAGREALPAWALT 60
 |||||
 QY 61 VYERAOYLHKAADIIERDKKEIATVLAKESIKAYNASVTEVVTADLIRYAAEGIELST 120
 DB |||||
 61 VYERAOYLHKAADIIERDKKEIATVLAKESIKAYNASVTEVVTADLIRYAAEGIELST 120
 |||||
 QY 121 SADEGGNDASTGHKLAVIRROPVGIIVLAIAPYPNVNLSSKIAIAPALIGGVNVEKPPT 180
 DB |||||
 121 SADEGGNDASTGHKLAVIRROPVGIIVLAIAPYPNVNLSSKIAIAPALIGGVNVEKPPT 180
 |||||
 QY 181 QGSVSGILVLAKAFAGLPAEAGLPAGVFNITGRSGEIGDIYVEHEEVNFINTGTPVQORIGK 240
 DB |||||
 181 QGSVSGILVLAKAFAGLPAEAGLPAGVFNITGRSGEIGDIYVEHEEVNFINTGTPVQORIGK 240
 |||||
 QY 241 LAGMRPIMLELGGKADGIVLADADLNNAAKQIVAGDYSGQRCTAIKRVLVVEVADEL 300
 DB |||||
 241 LAGMRPIMLELGGKADGIVLADADLNNAAKQIVAGDYSGQRCTAIKRVLVVEVADEL 300
 |||||
 QY 301 AEKISENVAKLSVGDPPDNATVTPVIDDMSADFIESLVVDAROKGAKELNEFKRGRLLT 360
 DB |||||
 301 AEKISENVAKLSVGDPPDNATVTPVIDDMSADFIESLVVDAROKGAKELNEFKRGRLLT 360
 |||||
 QY 361 PGLFDHVTLDMLKAWEPFGPILPIIRVKDAEEAVALANKSDFGLOSSVTRDFOKAFDI 420
 DB |||||
 361 PGLFDHVTLDMLKAWEPFGPILPIIRVKDAEEAVALANKSDFGLOSSVTRDFOKAFDI 420
 |||||
 QY 421 ANKLEVTGTHINNKTRGPNPFPLGLKSGAGVQGIYSIRSIEMTNVKSIVLDMK 475
 DB |||||
 421 ANKLEVTGTHINNKTRGPNPFPLGLKSGAGVQGIYSIRSIEMTNVKSIVLDMK 475
 |||||

RESULT 4

ABU01538
 ID ABU01538 standard; protein; 474 AA.

XX
 AC ABU01538;
 XX
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #1114.
 XX
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX

Streptococcus pneumoniae; type 4 strain.

WO200277021-A2.

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB0002163.

PR 27-MAR-2001; 2001GB-00007658.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Massignani V, Tettelin H, Fraser C;

DR WFI; 2003-040579/03.

DR N-PSDB; ABX06826.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX

Claim 1; SEQ ID NO 2228; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC A855454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX Sequence 474 AA;

Query Match 75.7%; Score 1812; DB 6; Length 474;
 Best Local Similarity 74.1%; Pred. No. 1.8e-142;
 Matches 349; Conservative 57; Mismatches 65; Indels 0; Gaps 0;
 QY 5 YQNYNGEWSKSVNQIEILSPDDSSIGFVPAMTREVDHAKAGREALPAWALTUYVER 64
 Db 4 YQNLVNGKWSQSEITVSPINQEEIGTVPAMTQTEADEAMQARALPAWALSAYER 63
 QY 65 AOTLHKADIIERDKEEITATVLAKEISKAYNASVTEVVRTADLIRYAAEEGRLSTSADE 124
 Db 64 AAYLHKTAAILERDKEEIGTILAKEVAKGKAAIGEVVVRTADLIRYAAEEGRLITQAME 123
 QY 125 GGMKDASTGHKLAVIRROPVGVILAIAPYNPVNLSSGSKIAPALIGNVVMPKPTQGSV 184
 Db 124 GGGFEATSKNKLAVVRREPVGIVLAIAPFNPVNLSSASKIAPALLAGNVVMPKPTQSSI 183
 QY 185 SGLVLAKAFAGAGPAGVNTITGSGEIGDYIVHEEVNFINTGSPVGORIGRLAGM 244
 Db 184 SGLLAKAFEEAGTIPAGVNTITGSGEIGDYIIIEHKEVNFINTGSPGIGRIGRLAGM 243
 QY 305 SENVAKLSVDPFDNATVTPVDDNSADFTESLWVDARQKAGKELNEFKRDLRLITGLF 364
 Db 304 QEEVSKLTVGDPFDNATVTPVDDNSADFTWGLIEDAQEKAQALTPIKREGNLIWPLVF 363
 QY 365 DHVTLDMKLAWEPFGPIPIIRVKDAEBAVAJANKSDFGLQSSVFTFROFQAFDIANKL 424
 Db 364 DQVTKMVKAWEEFFFGPIPIIRVASVEEAIAFANESEFGQLSSVFTNDFKKAFAIEKL 423
 QY 425 EVGTVHNNKTGPNPFPLGLKSGAGVQGGIRYSIEAMTNVKSIVLDMK 475
 Db 424 EVGTVHNNKTQGPDPNPFPLGVKSGAGVQGGIKYSIEAMTNVKSIVFDVK 474

RESULT 5
 ABP81589

ID ABP81589 standard; protein; 474 AA.
 XX
 AC ABP81589;
 XX
 DT 04-MAR-2003 (first entry)
 XX
 DE Streptococcus pneumoniae polypeptide SEQ ID NO 667.
 XX
 KW Streptococcus pneumoniae; infection; otitis media; antibacterial;
 XX diagnosis; gene therapy.
 XX
 OS Streptococcus pneumoniae.
 XX
 FN WO200283855-A2.
 XX
 PD 24-OCT-2002.
 XX
 PF 12-APR-2002; 2002WO-US011524.
 XX
 PR 16-APR-2001; 2001US-0283948P.
 PR 18-APR-2001; 2001US-0284443P.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;
 FI Wooters JL;
 XX
 DR WPI; 2003-093010/08.
 DR N-PSDB; AB242437.
 XX
 XX New Streptococcus pneumoniae polynucleotides, useful for treating or
 PT preventing *S. pneumoniae* infections, or non-systemic diseases, e.g.
 PT otitis media, which are induced or exacerbated by *S. pneumoniae*.
 XX
 Claim 42; Page 944-946; 1091pp; English.
 PS
 XX The invention relates to isolated polynucleotides (AB272147-AB242522) of
 CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate
 CC variant of the polynucleotide or a nucleic acid sequence 95% identical to
 CC one of the polynucleotides. The *S. pneumoniae* polynucleotides and encoded
 CC polypeptides (ABP81299-ABP81674) are useful for treating or preventing *S.*
 CC pneumoniae infections or non-systemic diseases, e.g. otitis media, which
 CC are induced or exacerbated by *S. pneumoniae*. These are also useful for
 CC detecting *S. pneumoniae* in a biological sample or diagnosing *S.*
 CC pneumoniae infection in a subject. The polynucleotides have antibacterial
 CC activity and are useful in gene therapy
 CC
 SQ Sequence 474 AA;

Query Match 75.7%; Score 1812; DB 6; Length 474;
 Best Local Similarity 74.1%; Pred. No. 1.8e-142;
 Matches 349; Conservative 57; Mismatches 65; Indels 0; Gaps 0;
 QY 5 YQNYNGEWSKSVNQIEILSPDDSSIGFVPAMTREVDHAKAGREALPAWALTUYVER 64
 Db 4 YQNLVNGKWSQSEITVSPINQEEIGTVPAMTQTEADEAMQARALPAWALSAYER 63
 QY 65 AOTLHKADIIERDKEEITATVLAKEISKAYNASVTEVVRTADLIRYAAEEGRLSTSADE 124
 Db 64 AAYLHKTAAILERDKEEIGTILAKEVAKGKAAIGEVVVRTADLIRYAAEEGRLITQAME 123
 QY 125 GGMKDASTGHKLAVIRROPVGVILAIAPYNPVNLSSGSKIAPALIGNVVMPKPTQGSV 184
 Db 124 GGGFEATSKNKLAVVRREPVGIVLAIAPFNPVNLSSASKIAPALLAGNVVMPKPTQSSI 183
 QY 185 SGLVLAKAFAGAGPAGVNTITGSGEIGDYIVHEEVNFINTGSPVGORIGRLAGM 244
 Db 184 SGLLAKAFEEAGTIPAGVNTITGSGEIGDYIIIEHKEVNFINTGSPGIGRIGRLAGM 243
 QY 245 RPTMLELGGKADAGIVLADADLNAAKQIVAGADYSGQRCTAIKRVLVVEEVADEIAEKI 304
 Db 244 RPTMLELGGKDAALVLEDADLEHAAKQIVAGAFSYSGQRCTAIKRVLVLESVADKATLL 303

CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 439 AA;

Query Match 71.0%; Score 1699; DB 4; Length 439;
 Best Local Similarity 74.5%; Pred. No. 4.3e-133;
 Matches 327; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

QY 37 MTRREVHAMKAGREALPAWALTYVYRAQYLHKAADIIRKKEIATVLAKEISKAYNA 96
 Db 1 MTQTEADAMQARAALPAWALSALERAAYLHKTAAILLERDKKICTILAKEVAKGIKA 60
 QY 97 SVTEVVRADLIRYAAEEGRISTSADEGGKMDASTGCHKLAVIRQPVGVILALAPYNP 156
 Db 61 AIGEVVRADLIRYAAEEGRITGQMEGGGFEAASKNLAVRRREPVGIVLALAPFNP 120
 QY 157 VNLGSKTAPALIGNVWFKPPTQGSVSLVLAKAFAGAGPAGVNTTIGRSGEIGDY 216
 Db 121 VNLGSKTAPALIGNVWFKPPTQGSVSLVLAKAFAGAGPAGVNTTIGRSGEIGDY 180
 QY 217 IVEHEVNFNFTGTPVQRIKGLAGMRPIMLELGGKAGIVLADLNAKQIVAGA 276
 Db 181 IIEHKEVNFNFTGTPIGRIGLAGMRPIMLELGGKDAALVLEADLEHAQIVAGA 240
 QY 277 YDYSQRCATKRVIAVEVADLAKEISNVAKLSVGGPDNPATVTPVIDNSADFTES 336
 Db 241 FYSQRCATKRVIVLESVADKJATLQEEVSKLTGVPFDNADITPVIDNSADFTWG 300
 QY 337 LVVDARQKAKELNEFKRDCRLLITPLGDFHVTLDMLAWEEPFGPILPIIRVKDAEEAVA 396
 Db 301 LIEDAQEKQAALTPKESGNLLMPVLDQVTKDMKVAWEFPGLPIIRVASVEEAIA 360
 QY 397 IANKSPGLQSSVTRDFOKAFDIANKLEVGTTHINNTKRGDPNFPFLGKSGAGVQG 456
 Db 361 FANESEFGLQSSVTFNDFKAFEAIEAKLEVGTTHINNTKRGDPNFPFLGKSGAGVQG 420
 QY 457 IRYSTEAMTNVKSIVLDMK 475
 Db 421 IRYSTEAMTNVKSIVFDVK 439

RESULT 9
 ABP27644
 ID ABP27644 standard; protein; 475 AA.
 XX
 AC ABP27644;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 4464.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 FN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-CB004789.
 XX
 PR 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;

PI Tettelin H;
 XX
 DR WPI: 2002-352536/38.
 DR N-PSDB; ABN68275.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX
 PS Claim 1; Page 3607; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 SQ Sequence 475 AA;

Query Match 68.6%; Score 1641; DB 5; Length 475;
 Best Local Similarity 65.7%; Pred. No. 3.4e-128;
 Matches 312; Conservative 76; Mismatches 87; Indels 0; Gaps 0;

QY 1 LITEYQNYNGEWKSSVNOQIEILSPIDSSLSGTFPAMTREEDVHAKGRALPAWALT 60
 Db 1 LAKQYNLVNGEKLSENEITTYAPATGEELGSPVAMTQAEVDVYASAKKALSDRRALS 60
 QY 61 VYERAOYLHKAADIIRDKKEIATVLAKEISKAYNASVTEVVRADLIRYAAEEGRIST 120
 Db 61 YVERAOYLHKAADIIRDKDAEIKGAILSKVEKAKHKAASEVIRTAELIYAAEEGLEMEG 120
 QY 121 SADEGGKMDASTGCHKLAVIRQPVGVILALAPYNPVNLGSKTAPALIGNVWFKPPT 180
 Db 121 EVLEGSGFEAASKKIALVRRREPVGIVLALSPFNYPVNLGSKTAPALIGNVWFKPPT 180
 QY 181 QGSVGLVILAKAFAPAGLPAGVNTTIGRSGEIGDYIVEHEEVEFNFTGTPVQRIK 240
 Db 181 QGSISGLLALAEAFAPAGLPAGVNTTIGRSGVIGDYIVEHEAVSFNFTGTPIGEGIK 240
 QY 241 LAGMRPIMLELGGKAGIVLADLNAKQIVAGAYDYGORCTAIKRVLVVEVADEL 300
 Db 241 LAGMRPIMLELGGKDAISALVEDADLALAKVIVAGAGYSGQRCTAVKRVLMKVDQL 300
 QY 301 AEKISENVAKLSVGGPDNPATVTPVIDNSADFTESLVVDARQKAGKELNFKDGRLLT 360
 Db 301 AAEIKLVEKLSVGMPEDDADITPLIDTSAADFVEGLIKDATKGTALTAFAFNEGGLIS 360
 QY 361 PGLFDHVTLDMLAWEEPFGPILPIIRVKDAEEVAIAANKSDFGLOSSTVTRDFQKAFDI 420
 Db 361 PVLFDHVTLDMLAWEEPFGPILPIIRVTVEEAIKISNESEYGLQASIFTNFPKAFGI 420
 QY 421 ANKLEVGTTHINNTKRGDPNFPFLGKSGAGVQGIKRVISIEAMTNVKSIVLDMK 475
 Db 421 ABQLEVGTVHLNNKIQRTGTDNFPFLGAKKSGAGVQGVKYSIEAMTNVKSIVFDIQ 475

RESULT 10
 ABU46701
 ID ABU46701 standard; protein; 475 AA.
 XX
 AC ABU46701;

Best Local Similarity 56.7%; Pred. No. 4.1e-107;
Matches 268; Conservative 90; Mismatches 112; Indels 3; Gaps 2;

QY 5 YONYUNGEMKSSV--NQIEILSPIDSSLSGFPVPMTRREVDHAKAGREALPAAALTVY 62
DB 12 YKNFYGQWNSKTNFTFIEISSPIDSLGKVPKSEEDVLAIAKAKAQSHNEIPIN 71
QY 63 ERAQYLHKAADIIRKDEKEIATVLAKEISKAYNAVTEVVTADLRVAABEGIRLTS 122
DB 72 EKATILLKAAHILDEKABEADIIMTKELAKDKKSSISEVTRTADYIRFSADTAKNMVGET 131
QY 123 DEGGKMDATGCHKLAVIRROPVGVILAIAPYVNVNLSGSKIAPALIGNVVMFKPPTQG 182
DB 132 LFGDSFNGSKGISIVNRVPLGVLAISFPFNVNLSGSKIAPALMAGNSVVLKPTQG 191
QY 183 SVSGLVLAKAFAPAGLPAAGVNTITGRSEIGDYIVHEEVNFINTGSTVPGQIRIGLA 242
DB 192 SISALHLVKFEAGLPGVLTITGKSEIGDYIVSHPMIDFINTGSTVPGKRIHIS 251
QY 243 GMRPIMLEGGKAGIVLADLNAKQIVAGAYDSGQRTAKIRVLVVEVADELA 302
DB 252 VKMFMELGGKDAIVLEDDADLAAKNIVSGAYSYSGQRTAVKRLVLEETADTIVE 311
QY 303 KISENVAKLSVGDPF-DNATVTPVIDNSADFIESLVVDARQKAGKELNEKRDGRLLTP 361
DB 312 KVKERSVSLKIGNPLKEQVDIVLDDKADYVEALMBEANNKGATLVGGKREGNLIYP 371
QY 362 GLFDHVTLDMLKAWEEPFGPIPIIRYKDAEVAIAKNSDFGLQSSVFTDRDFOKADIA 421
DB 372 TLFNDVTLDMRLAWEEPAPILPIIRYKDVDEAIRANESYGLQASIFTKKINKAPYA 431
QY 422 NKLEVGTVHNNKTRGDPNPFPLGLKSGAGVQIRYSIEAMTNVKSIVLDM 474
DB 432 BKLEVGTVQNNKPERGDPHPFPGTIGKASGLGTQIKYSIEAMSRPKAIVLNI 484

RESULT 13
ABU24149
ID ABU24149 standard; protein; 482 AA.
AC ABU24149;
XX
XX 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #9676.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Clostridium acetobutylicum.
XX
FN WC200277183-A2.
XX
PD 03-OCT-2002.
XX
XX 21-MAR-2002; 2002MO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI: 2003-029926/02.
DR N-PSDB; ACA28019.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 52073; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 482 AA;
Query Match 57.8%; Score 1383; DB 6; Length 482;
Best Local Similarity 57.0%; Pred. No. 1.2e-106;
Matches 269; Conservative 86; Mismatches 115; Indels 2; Gaps 2;

QY 5 YONYUNGEM-KSSVQ-LEILSPIDSSLSGFPVPMTRREVDHAKAGREALPAAALTVY 62
DB 11 YKNFYGQWNSKTNFTFIEISSPIDSLGKVPKSEEDVLAIAKAKAQSHNEIPIN 70
QY 63 ERAQYLHKAADIIRKDEKEIATVLAKEISKAYNAVTEVVTADLRVAABEGIRLTS 122
DB 71 ERARIMRKAADILDDNAEYIAKILSNELAKOLKSSLSSEVKETADFRFTANEGTHMEGA 130
QY 123 DEGGKMDATGCHKLAVIRROPVGVILAIAPYVNVNLSGSKIAPALIGNVVMFKPPTQG 182
DB 131 INSDNFFGSKKDKLSILVERVPLGVLAISFPFNVNLSGSKVAPALIGNSVVLKPTTG 190
QY 183 SVSGLVLAKAFAPAGLPAAGVNTITGRSEIGDYIVHEEVNFINTGSTVPGQIRIGLA 242
DB 191 ALSALHUAETFNAGLPAAGVNTVTKGSEIGDYILTHEEVNFINTGSSAVGKHISKIA 250
QY 243 GMRPIMLEGGKAGIVLADLNAKQIVAGAYDSGQRTAKIRVLVVEVADELA 302
DB 251 GMTFVWLELGGKDAIVLEDDADLAAKNIVSGAYSYSGQRTAVKRLVMDKVADELVE 310
QY 303 KISENVAKLSVGDPFDPNATVTPVIDNSADFIESLVVDARQKAGKELNEKRDGRLLTPG 362
DB 311 LVTKVKELKVGKNPFDDVTITPLDINKAADYVQTLDDAIEKAGATLVGNKREKNLMYPT 370
QY 363 LFDHVTLDMLKAWEEPFGPIPIIRYKDAEVAIAKNSDFGLQSSVFTDRDFOKADIA 422
DB 371 LFDNVTADMRLAWEEPFGVPLPIIRYKSDVEAIEANRSEIGLQSAFTENHDAFTIAN 430
QY 423 KLEVGTVHNNKTRGDPNPFPLGLKSGAGVQIRYSIEAMTNVKSIVLDM 474
DB 431 KLDVGTVQNNKPERGDPHPFPLGLTSSGKSGVQIRYSIEAMTRHKSIVLNI 482

RESULT 14
AAG40912
ID AAG40912 standard; protein; 498 AA.
XX AC AAG40912;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 50828.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139859P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.

Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134321P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0141287P.
PR 30-JUN-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144311P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149910P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.

PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155133P.
PR	23-SEP-1999;	99US-0155488P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	18-OCT-1999;	99US-0159638P.
PR	21-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	22-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match 48.8%; Score 1168.5; DB 3; Length 496;
Best Local Similarity 50.3%; Pred. No. 9.9e-89;
Matches 240; Conservative 84; Mismatches 142; Indels 11; Gaps 5;

QY	5	YONVNGEMK--SSVNOQLTEILSPIDDSLSLGFVPAPMTREEDVDHAKAGREALPAWALTYV 62
DB	16	YKYAGGEWKTSSSKSVAINMPATRTQYKQVQACTQEEVNAVWELAKSAQSWAKTPLW 75
QY	63	ERAGYLHKAADIIERKEEIAIVLAKEISKAYNAVSTEVVTRTADLIRYAAEEGIRLSTSA 122
DB	76	KRAELLHKAALILKONKAPAESLIVKEIAKPAKDSVTEVRSGLISYCAEGVRI---L 132
QY	123	DEGKM--DASTGH---KLAVIRQVGVIAIPYVPYVNSGGKIAPALIGNVVMXK 177
DB	133	GEKFLISDFPNDRTYCLTSKIPLGVLAIPFPYVYVNAVSKIAPIALIGNSVLK 192
QY	178	PPTGSGVLAKAFAPAGVNTITGSGEIGDYIVHEEVNFINTGTFPVQR 237
DB	193	PPTGASVCLHWCHFHUAGPGLISCTGKGSEIGDFTMPAVNCISFTGG-DTGIS 251
QY	238	IGKLAGMRPTIMLEGGKAGIVLADADLNAAKQIVAGAYDYGSGQRTAKRVLVVEVA 297
DB	252	ISKAGMIPLQWELGKDACIVLDDDLVASNIKGGFSYSGQRCTAVKVLVWESVA 311
QY	298	DELAEKISENVAKLSVGDPFDNATVPTVIDDNSADFTESLVDAKQKAKELNEFFRDR 357
DB	312	DELVEKPKAKVAKLVGTPPEENSIDITAVVSESSANFTIEGLVMDAKEKATFCQBYKREGN 371
QY	358	LLTGLFDHVTLDKWLAEPPFGPILPIIRVKDAEBAVAJANKSDFGI-QSSVFFTRDFQKA 417

DB	372	LIWPLLLDNVRPDMRIAWEEPFGVPVLRINSVEEGINHCNASNFGLOGCVFTKDINKA 431
QY	418	FDIANKLEVGTVEHNNKTRGPDNFPPLGKSGAGVGGIRYSTEAMTVKSVILDM 474
DB	432	ILISDAMETGTQVINSAPARGPDHFPQGLKDSGSGQGVNTNSINLMTKVKTVINL 488

RESULT 18
AAG05753
ID AAG05753 standard; protein; 496 AA.
XX
AC AAG05753;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 2276.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
PN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125780P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132487P.
PR 14-MAY-1999; 99US-0132487P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 13-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140355P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140931P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142277P.
 PR 13-JUL-1999; 99US-0143342P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144684P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.

PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159594P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 48.7%; Score 1164.5; DB 3; Length 496;
 Best Local Similarity 50.1%; Pred. No. 2.le-88;
 Matches 239; Conservative 84; Mismatches 143; Indels 11; Gaps 5;
 Qy 5 YONYNGEWK--SSVNCIRILSPIDSSIGFVPMTRFEEVDHAKKAGREALPAWALTY 62
 Db 16 KYIADGEWTSSTSGKSVAILNPATRKTYKQVQCTQEBEWNWELAKSAQKSWAKTPIW 75

PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148664P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.

PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 47.2%; Score 1128.5; DB 3; Length 461;
 Best Local Similarity 51.6%; Pred. No. 1.9e-85;
 Matches 230; Conservative 78; Mismatches 129; Indels 9; Gaps 4;

Qy 34 VPAMTREVDMHAKAGREALPAWAALTYVERAQLYHKAADIIRKDEBIATVLAKEISKA 93
 Db 12 VQACTOEYVNAVWELAKSAQKSWAKTFLMKRAELLHKAAILKDNKAPMAESLVKETAKP 71
 Qy 94 YNASVTEVURTADLIRYAAEEGIRLSTSDSGKK--DASTGH--KLAVIRROPVGIVL 148
 Db 72 AKOSVTEVVRSGDLISYCABEGVRI---LGEKFLSDSPGNDRTKCYLTSKIPJGVVL 128
 Qy 149 ALAPYNYPNVLSGSKIAPALIGNVVMFKPTQGSVGLVLAKAFAEAGLPAGVFNITIG 208
 Db 129 AIPPNYPNVLAISKIAPALIAGNSLVLPKPTQGVSLHVMVHCFLAGFPKGLISCIITG 188
 Qy 209 RGSEIGDYIHEHEVNFINTGSTPVGQRIKLAGMPIMLELGKDAGIVLADADLNA 268
 Db 189 KGSEIGDFTLHPAVNCISFTGG-DTGISISKAGMIPLQMEILGGKDACIVLDDADLVL 247
 Qy 269 AKQIVAGYDYGORCTAIRVLVVEEVADELAEKISENVAKLSVGDPPDNATVTPVID 328
 Db 248 ASNIIRGSGFSYGRCTAVKVLWVESVADELVEKVKAKVLITVGPPEENSDITAVSE 307
 Qy 329 NSADFIESLVVDARQKAGKELNEFKDGRLLTPGLFDHVTLMKLAWEPEPGILPIRV 388
 Db 308 SSANFIEGLVMDAKEKATCQEKYKEGNLIWPLLNDVNPDMRIAWEEPFGVPVPLRI 367
 Qy 389 KDAEEAVAIANKSDFGLQSSVTRDFQKAFDIANKLEVGVTHNNKTKGRDNPFFGLK 448
 Db 368 NSVEEGINHCNASNFGQGCVFTKDKAILISDAMETGTQINSAPARGEDHPPFQGLK 427
 Qy 449 GSGAGVQIGIRYSLEAMTNVKSIVLDM 474
 Db 428 DSGIGSQGVINSINLMTRKVTINL 453

RESULT 20

AAG49551
 ID AAG49551 standard; protein; 461 AA.

XX AAG49551;

AC AAG49551;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 62698.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.

XX Arabidopsis thaliana.

XX EF1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131443P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132488P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.

05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158363P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159332P.
PR 14-OCT-1999; 99US-0159333P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 47.28; Score 1128.5; DB 3; Length 461;
Best Local Similarity 51.66; Pred. No. 1.9e-85;
Matches 230; Conservative 78; Mismatches 129; Indels 9; Gaps 4;
QY 34 VPATREEDVHAMKAGREALPAWALTVYERAQVYLHKAADIERKEEITATVLAKESKA 93
Db 12 VQACTQEEVNAVVELASQAKWAKTLPKRAELHKAALIKONKAPMAESVKEIAXP 71
QY 94 YNASVTEVVRADLIIRYAAEEGIRLSTGADGGRM--DASTGH--KLAVIRPQPVGVL 148
Db 72 AKDSVTEVVRSGDLISYCAEEGVRI--LGEGRFLLSDSFGPNDRTKVLTSKIPGVUL 128
QY 149 ATAPVNYPNLSGSKTAPALIGNVVMFKPTQGSVGLVLAKAPAEAGLPAGVENTITG 208
Db 129 AIPPNYPNVNLAVSKTAPALAGNSLVLPKPTQGAVSCLHMVHCFHLAGFPKGLISCITG 188
QY 209 RGEIGDIVIVEHEVNFNFTGSTPVGQRIKLAGMRPIMLEGGKDGAGIVLADADLNA 268
Db 189 KGEIGDPLTWHPAVNCISFTGG-DTGISISKAGMIPLOMELGGKDACIVLDADLDLV 247
QY 269 AKQIVAGAYDSGQRCATKRVLVEEVADELAKEISENVAKLSVGDPFNATVTPVIDD 328
Db 248 ASNIIKGGFSYSGQRCATKRVLVMSVADELVEKVKAKVAKLTGVPPEENSIDTAVVSE 307
QY 329 NSADFTBSLVVDARQKAGNELFEKSDGRLITPCLFDHYTLDMLKAMEEFGFIPITRV 388
Db 308 SSANFTGLVWDKAEKANTFCQYKREGNLWPLLDNVRPDMRTAWEEFGFVGVVPLRI 367
QY 389 KDAEEAVAIANKSDFGLQSSVTRDFQKAFDIANKLEVGTVHNNKTGRGDFNFPFLGLK 448
Db 368 NSVEEGINHCNASNFGVGGCVFTKQINKAIIISDMETGTQVINSAPARGDFHFPQGLK 427
QY 449 GSGAGVQIRISISAMTNVKSIVLDM 474
Db 428 DSGIGSQGVTSNINLMTKVTIVNL 453

ID AAG49552 standard; protein; 438 AA.
AC AAG49552;
XX 18-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 62699.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX EF1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 17-JUN-1999; 99US-0139452P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.

PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 46.1%; Score 1103.5; DB 3; Length 438;
Best Local Similarity 51.8%; Pred. No. 2.2e-83;
Matches 224; Conservative 76; Mismatches 125; Indels 9; Gaps 4;
QY 46 MKAGREALPAWALTIVYERAQYLVKAAADIIERDKERIATVLAKESIKAYNASVTEWRTA 105
Db 1 MELAKSAQSWAKTFLWKAEALLHKAAILKONKAPMAESLVKEIAPKADSVTEVRS 60
QY 106 DIIRYAAEEGIRLSADEGKX--DASTGH--KIAVIRQPVGIVLAIPYVNVLS 160
Db 61 DLISYCAEEGVRI---LGEKXLLSDSPGNDRTKYCLTSKIPLGVLVAIPPNVFNLA 117
QY 161 GSKIALPALIGNVMFKPTQGSVGLVAKAPAEAGLPAGVFNITGRSGSIGDIYVEH 220
Db 118 VSKIALPALIGNSLVLKPTQGAVSCLHVFCHLAGFPKGLISCIITGKSGSIGDFTMH 177
QY 221 EENVNFTGTSTPVQORIGKLAGMRPIMLELGGKDGAGIVLADADLONAAKQIVAGAYDS 280

Db 178 PAVNCISFTGG-DTGISISKAGMPLQMGELGGKACIVLDDADLDIVASNIKGFYS 236
QY 281 GORCTATKRVLVVEVADLAETKISENVAKLSVGGPFEDNATVTPVIDDASADFIISLWVD 340
Db 237 GORCTAVKVLWVESVADLEKVKAKAKVAKLIVGPPENSDDITAVSESSAFIEGLWVD 296
QY 341 ARQKAKELNEFKRDLTLTGLFDHVTLDMLAMEEPFGELPIIRVKDAEEVAIAANK 400
Db 297 AKKKGATFCQYKKEGKLNPLLDNVRPDMRIAMEEPPGVPVLRINSVEEGINHCNA 356
QY 401 SDFGLQSVTRDFQAFDIANKLEVGTVINNKTRGPDNFPPLGLKSGAGVGTIRYS 460
Db 357 SNFGLQCVTFKDIKAILISDAMETGTQINSAPARGPDHFPQGLKSGIGSQGVINS 416
QY 461 IEAMTVKKSIVLDM 474
Db 417 INLMTKVTIVNI 430

RESULT 22
AAG32609
ID AAG32609 standard; protein; 438 AA.
XX AC AAG32609;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 39373.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX FN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129645P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144864P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

PR	03-AUG-1999;	93US-01470389;
PR	04-AUG-1999;	93US-01472049;
PR	04-AUG-1999;	93US-01473029;
PR	05-AUG-1999;	93US-01471929;
PR	05-AUG-1999;	93US-01472609;
PR	06-AUG-1999;	93US-01473039;
PR	06-AUG-1999;	93US-01474169;
PR	06-AUG-1999;	93US-01474939;
PR	09-AUG-1999;	93US-01479359;
PR	10-AUG-1999;	93US-01481719;
PR	11-AUG-1999;	93US-01483199;
PR	12-AUG-1999;	93US-01483419;
PR	13-AUG-1999;	93US-01485659;
PR	13-AUG-1999;	93US-01486689;
PR	16-AUG-1999;	93US-01493689;
PR	17-AUG-1999;	93US-01491759;
PR	18-AUG-1999;	93US-01494269;
PR	20-AUG-1999;	93US-01497229;
PR	20-AUG-1999;	93US-01497239;
PR	20-AUG-1999;	93US-01499299;
PR	23-AUG-1999;	93US-01493029;
PR	23-AUG-1999;	93US-01499309;
PR	25-AUG-1999;	93US-01505669;
PR	26-AUG-1999;	93US-01508849;
PR	27-AUG-1999;	93US-01508549;
PR	27-AUG-1999;	93US-01510669;
PR	27-AUG-1999;	93US-01510809;
PR	30-AUG-1999;	93US-01513039;
PR	31-AUG-1999;	93US-01513489;
PR	01-SEP-1999;	93US-01519309;
PR	07-SEP-1999;	93US-01523639;
PR	10-SEP-1999;	93US-01530709;
PR	10-SEP-1999;	93US-01537589;
PR	15-SEP-1999;	93US-01540189;
PR	16-SEP-1999;	93US-01540399;
PR	20-SEP-1999;	93US-01547799;
PR	22-SEP-1999;	93US-01551339;
PR	23-SEP-1999;	93US-01554869;
PR	24-SEP-1999;	93US-01556599;
PR	28-SEP-1999;	93US-01564589;
PR	29-SEP-1999;	93US-01565969;
PR	04-OCT-1999;	93US-01571179;
PR	05-OCT-1999;	93US-01577539;
PR	06-OCT-1999;	93US-01578659;
PR	07-OCT-1999;	93US-01580299;
PR	08-OCT-1999;	93US-01582329;
PR	12-OCT-1999;	93US-01583699;
PR	13-OCT-1999;	93US-01592939;
PR	13-OCT-1999;	93US-01592939;
PR	14-OCT-1999;	93US-01593299;
PR	14-OCT-1999;	93US-01593339;
PR	14-OCT-1999;	93US-01593339;
PR	14-OCT-1999;	93US-01596379;
PR	14-OCT-1999;	93US-01596389;
PR	18-OCT-1999;	93US-01595849;
PR	21-OCT-1999;	93US-01607179;
PR	21-OCT-1999;	93US-01607469;
PR	21-OCT-1999;	93US-01607689;
PR	21-OCT-1999;	93US-01607709;
PR	21-OCT-1999;	93US-01608149;
PR	22-OCT-1999;	93US-01608159;
PR	22-OCT-1999;	93US-01609809;
PR	22-OCT-1999;	93US-01609819;
PR	22-OCT-1999;	93US-01609899;
PR	25-OCT-1999;	93US-01614049;
PR	25-OCT-1999;	93US-01614059;
PR	25-OCT-1999;	93US-01614069;
PR	26-OCT-1999;	93US-01613599;
PR	26-OCT-1999;	93US-01613609;
PR	26-OCT-1999;	93US-01613619;
PR	28-OCT-1999;	93US-01619209;
PR	28-OCT-1999;	93US-01619929;

PR	28-OCT-1999;		99US-0161993P.	
PR	29-OCT-1999;		99US-0162142P.	
Query Match 46.1%; Score 1103.5; DB 3; Length 438;				
Best Local Similarity 51.6%; Pred. No. 2.2e-83;				
Matches 224; Conservative 76; Mismatches 125; Indels 9; Gaps 4				
QY	46	MKAGREAPDAWAALTVVERAQYLHKAAADIIRDEKEIATVLAKEISKAYNAVSVEVVRTA	105	
Dd	1	MELAKSAQSKWAKTPLWKRAELLHKAAILDKNKPMAESLVEIAKPAKDSVTEVVRSG	60	
QY	106	DLTRYAREEGRILTSADGGKM--DASTGH---KLAVIRQPVGIVLAIPNYPNLS	160	
Dd	61	DLISYCABEGVRI--IGBGKFLUSDSPFGNDRTKYCTSKIPLGVVLAIPPFNEYVNLA	117	
QY	161	GSKIAPALIGNVVMFKPTQGSVGLVLAKAFAEAGLPAGVFNTITGSGSIGDYIVEH	220	
Dd	118	VSKIAPALIAGNSLVLPKPTQGANVSLHWVCHFLAGFPKGLISCITGKGSEIGDFLTMH	177	
QY	221	EENVFINFTGSTPYGQRIGIKLAGMRPIMLBGLGKDAGIVLADADLNAAKQIVAGAYDYS	280	
Dd	178	PAVNCISFTGG-DTGISISKAGMIPLQMBELGGKDACIVLDDADDLVASNIKGGFSYS	236	
QY	281	GQRCTAIRKVLWEVADELAKISENVAKLSVGDPFNATVTVTDIDNSADFIESLWVD	340	
Dd	237	GQRCTAVKVLMVESVADELVEKKVAKYKLTVPGPENSDITAVVSSSANFIIEGLVMD	296	
QY	341	AROKGAKEINFEKDGRLTLTPGLFDHVTLDMKLAWBPFGPILPIRVKDAEAVALANK	400	
Dd	297	AKSGKATFCQBYKEEKNLIWPLLNDNRPMRIAMEEPFGPVVEVKINSVEEINHENA	356	
QY	401	SDFLGQSFFTRDFQKAFDIANKLEVTGTHNNKTGRGPDNFPPFLGIKSGAGVQGIRYS	460	
Dd	357	SNFGLOCVFYKDKINKALLISDAMETGTVOINSAPARGEDHFPEQGLKDSGGISGVVNS	416	
QY	461	IEAWTNVKSIVLDM 474		
Dd	417	INLMTKVKTVINL 430		
RESULT 23				
ID	AAG40913			
AC	AAG40913 standard; protein; 438 AA.			
XX	AAG40913;			
DT	18-OCT-2000 (first entry)			
XX				
XX				
XX				
KX	Zea mays protein fragment SEQ ID NO: 50829.			
KW	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
KW	termination sequence; corn.			
XX				
XX	Zea mays subsp. mays.			
XX				
XX	EP1033405-A2.			
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000; 2000EP-00301439.			
PR	25-FEB-1999;	99US-0121825P.	XX	
PR	05-MAR-1999;	99US-0123180P.	XX	
PR	09-MAR-1999;	99US-0123348P.	XX	
PR	23-MAR-1999;	99US-0125788P.	XX	
PR	25-MAR-1999;	99US-0126264P.	XX	
PR	29-MAR-1999;	99US-0126785P.	XX	
PR	01-APR-1999;	99US-0127462P.	XX	
PR	06-APR-1999;	99US-0128234P.	XX	
PR	08-APR-1999;	99US-0128714P.	XX	
PR	16-APR-1999;	99US-0129845P.	XX	
PR	19-APR-1999;	99US-0130077P.	XX	

PR	21-APR-1999;	99US-0130449P.	PR	20-JUL-1999;	99US-0144352P.
PR	23-APR-1999;	99US-0130510P.	PR	20-JUL-1999;	99US-0144632P.
PR	23-APR-1999;	99US-0130891P.	PR	20-JUL-1999;	99US-0144884P.
PR	28-APR-1999;	99US-0131449P.	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	99US-0132048P.	PR	21-JUL-1999;	99US-0145086P.
PR	30-APR-1999;	99US-0132407P.	PR	21-JUL-1999;	99US-0145088P.
PR	04-MAY-1999;	99US-0132484P.	PR	22-JUL-1999;	99US-0145085P.
PR	05-MAY-1999;	99US-0132485P.	PR	22-JUL-1999;	99US-0145087P.
PR	06-MAY-1999;	99US-0132486P.	PR	22-JUL-1999;	99US-0145089P.
PR	06-MAY-1999;	99US-0132487P.	PR	22-JUL-1999;	99US-0145132P.
PR	07-MAY-1999;	99US-0132863P.	PR	23-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	99US-0134256P.	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	99US-0134218P.	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	99US-0134219P.	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	99US-0134221P.	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	99US-0134370P.	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	99US-0134768P.	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	99US-0134941P.	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	99US-0135424P.	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	99US-0135533P.	PR	02-AUG-1999;	99US-0146388P.
PR	24-MAY-1999;	99US-0135629P.	PR	02-AUG-1999;	99US-0146389P.
PR	25-MAY-1999;	99US-0136021P.	PR	03-AUG-1999;	99US-0147038P.
PR	27-MAY-1999;	99US-0136392P.	PR	04-AUG-1999;	99US-0147204P.
PR	28-MAY-1999;	99US-0136782P.	PR	04-AUG-1999;	99US-0147302P.
PR	01-JUN-1999;	99US-0137222P.	PR	05-AUG-1999;	99US-0147192P.
PR	03-JUN-1999;	99US-0137528P.	PR	05-AUG-1999;	99US-0147260P.
PR	04-JUN-1999;	99US-0137502P.	PR	06-AUG-1999;	99US-0147303P.
PR	07-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	18-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	99US-0139899P.	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140353P.	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0152363P.
PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0142154P.	PR	22-SEP-1999;	99US-0154779P.
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-015753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158636P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159283P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.

PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 46.1%; Score 1102.5; DB 3; Length 438;		
Best Local Similarity 50.6%; Pred. No. 2.7e-83;		
Matches 220; Conservative 78; Mismatches 126; Indels 11; Gaps 4;		
Qy	46	MKAGREALPAWALVYERQYLKKAADIHERDKEEATYLAKEISKAYNASVTEVURTA 105
Db	1	MDAKVAQAWARTPLWKRAELHKAALILKEHKAPAECLVKEIAKPAKDAVSEVRSG 60
Qy	106	DLRYAABEGIRLSADEGKM---DASTG---HKLAIRPQPVGIVLAIAPNYFVNL 159
Db	61	DLVSYTAEGSVRIIE---GKLVSDFSGNERNKYCLSSKTLPLGVLAIPPNYFVNL 116
Qy	160	SGSKIAPALIGVVWVWPKPTQSGVGLVLAKAFAEAGLPAGVNTITGRSGEIGDIYVE 219
Db	117	AVSKIGPALIAGNALVLKPTQGAVALHMHVCHFLAGFPKGLISCVTGRSGEIGDFTM 176
Qy	220	HEEVNFTNTGTSTPGVQRIKGLAGMRPIMLELGGKDGAGIVLADADLNAKQIVAGAYDY 279
Db	177	HPGVNCISTGG-DTGTAISKAGVPLQNELGGKDACIVLEDDDLVSNIVKGGFY 235
Qy	280	SGQCTAIRKVLVVEEADLAKISENVAKLSVGDFDNATVPTVIDNSADPIESLVV 339
Db	236	SGQCTAVKVLIMESADAVQVNAKLKVGPPEDSDTTPWTSSANFIEGLVW 295
Qy	340	DARQKAKELNEFKEDGELLTPGLFDHVTLDKLANEPEPGPLPIIRVKDAEVAIAN 399
Db	296	DAKEGATTCQYREBGLNPLLLDHPVPMRIAWEEPFGVLPVIRINSVEEGIHCCN 355
Qy	400	KSDPGLQSSVFRDFQAKFOITANKLEVGVTHINNKTRGPDNFPPLGLKSGAGVQIGRY 459
Db	356	ASNFGQGCIFTRDKNALLISDAMETGTQVNSAPARGDPHPFQGLKDSGLSGQIIN 415
Qy	460	SIEMNVKSVILDM 474
Db	416	SINMTVKVSTVINL 430

RESULT 24
AAG05754
ID AAG05754 standard; protein; 438 AA.
XX
XX AAG05754;
AC
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 2277.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW

KW	termination sequence.
XX	Arabidopsis thaliana.
OS	EP1033405-A2.
PN	06-SEP-2000.
XX	25-FEB-2000; 2000EP-00301439.
PD	25-FEB-1999; 99US-0121825P.
XX	05-MAR-1999; 99US-0123180P.
PR	09-MAR-1999; 99US-0123548P.
PR	23-MAR-1999; 99US-0125780P.
PR	25-MAR-1999; 99US-0126264P.
PR	29-MAR-1999; 99US-0126785P.
PR	01-APR-1999; 99US-0127462P.
PR	06-APR-1999; 99US-0128234P.
PR	08-APR-1999; 99US-0128714P.
PR	16-APR-1999; 99US-0129845P.
PR	19-APR-1999; 99US-0130077P.
PR	21-APR-1999; 99US-0130449P.
PR	23-APR-1999; 99US-0130510P.
PR	23-APR-1999; 99US-0130891P.
PR	28-APR-1999; 99US-0131449P.
PR	30-APR-1999; 99US-0132048P.
PR	30-APR-1999; 99US-0132407P.
PR	04-MAY-1999; 99US-0132484P.
PR	05-MAY-1999; 99US-0132485P.
PR	06-MAY-1999; 99US-0132486P.
PR	06-MAY-1999; 99US-0132487P.
PR	07-MAY-1999; 99US-0132863P.
PR	11-MAY-1999; 99US-0134256P.
PR	14-MAY-1999; 99US-0134218P.
PR	14-MAY-1999; 99US-0134219P.
PR	14-MAY-1999; 99US-0134221P.
PR	14-MAY-1999; 99US-0134370P.
PR	18-MAY-1999; 99US-0134768P.
PR	19-MAY-1999; 99US-0134941P.
PR	20-MAY-1999; 99US-0135124P.
PR	21-MAY-1999; 99US-0135353P.
PR	24-MAY-1999; 99US-0135629P.
PR	25-MAY-1999; 99US-0136021P.
PR	27-MAY-1999; 99US-0136392P.
PR	28-MAY-1999; 99US-0136782P.
PR	01-JUN-1999; 99US-0137222P.
PR	03-JUN-1999; 99US-0137528P.
PR	04-JUN-1999; 99US-0137502P.
PR	07-JUN-1999; 99US-0137724P.
PR	08-JUN-1999; 99US-0138094P.
PR	10-JUN-1999; 99US-0138540P.
PR	10-JUN-1999; 99US-0138847P.
PR	14-JUN-1999; 99US-0139119P.
PR	16-JUN-1999; 99US-0139452P.
PR	16-JUN-1999; 99US-0139453P.
PR	17-JUN-1999; 99US-0139492P.
PR	18-JUN-1999; 99US-0139454P.
PR	18-JUN-1999; 99US-0139455P.
PR	18-JUN-1999; 99US-0139456P.
PR	18-JUN-1999; 99US-0139457P.
PR	18-JUN-1999; 99US-0139458P.
PR	18-JUN-1999; 99US-0139459P.
PR	18-JUN-1999; 99US-0139460P.
PR	18-JUN-1999; 99US-0139461P.
PR	18-JUN-1999; 99US-0139462P.
PR	18-JUN-1999; 99US-0139463P.
PR	18-JUN-1999; 99US-0139750P.
PR	18-JUN-1999; 99US-0139763P.
PR	21-JUN-1999; 99US-0139817P.
PR	22-JUN-1999; 99US-0139899P.
PR	23-JUN-1999; 99US-0140353P.
PR	23-JUN-1999; 99US-0140354P.
PR	24-JUN-1999; 99US-0140695P.

```

PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144864P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145113P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.

```

```

PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```

```

Query Match 45.9%; Score 1099.5; DB 3; Length 438;
Best Local Similarity 51.4%; Pred. No. 4.8e-83;
Matches 223; Conservative 76; Mismatches 126; Indels 9; Gaps 4;

Qy 46 MKAGREALPANAALTYYERAQYLHKAADIIBRDKEETATVLAKEISKAYNASVTEVVRTA 105
Db 1 MELAKSAQKSWAKTPLWKRTTELLHKAALIKONKAPWAESLVKEIAKPAKDSVTEVVRSG 60

Qy 106 DLIRYAABEGIRLSTSDAEGKM--DASTGH---KLAVIRPQVGLVLAIPNYPNLS 160
Db 61 DLISYCAEEGVRI--LGEKGKLLSDSPGNDRTKYCLTSKPLGVLAIAPPNYPNLA 117

Qy 161 GSKTAPALIGNVNMFPTQGSVGLVLAKAFPAEAGLPAGVFNITIGRSEIGDYVEH 220
Db 118 VSKTAPALTAGNSLVLPKPTQGSVCLHWHCFHLAGPKGLISCTIGKSGEIGDPLTWH 177

Qy 221 EYVNFNFTGTPVGQRIGKLAGMRPIMLELGGKAGIVLADLDNAAKQIVAGAYDS 280
Db 178 PAVNCISFTGG-DTGISISKAGMPLQWELGGKDACIVLDDADLDLVASNIKGGFSYS 236

Qy 281 GQCTAIKEVLVVEVADELAEKISENVAKLSDGDFDNATVTPVIDDONSADFTESLVVD 340
Db 237 GQCTAVKVLVWESVADELVEKVAKVAKLTGVPPEENS DITAVVSESSANFTLEGLVMD 296

Qy 341 AROKGAKELNEFKDGRLLPGLFDHVTLDMLKAWPEPFGPILPIIRVKDAEBAVAIAK 400
Db 297 AKKGATFCQEKREGNLIWPLLDNVRPDMRIANEEPPGVPVPLRINSVEEGINHCNA 356

```

Qy 401 SDFGLSSVTFDFQKAFDIANKLEVTGTVHNNKTGRPDNFFFLGLKSGAGVQIRYS 460
Db 357 SNFGLQCVFTDIKAILISDAMETGTQINSAPARGPDHFFQGLKXDSGSGQVTS 416
Qy 461 IEAMTVKSVILDM 474
Db 417 INLTKVKTIVNL 430

RESULT 25
AAG05755
ID AAG05755 standard; protein, 402 AA.
XX AAG05755;
AC AAG05755;
XX
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 2278.
DE
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141642P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.

PR	11-AUG-1999,	99US-01483113P,
PR	12-AUG-1999,	99US-01483411P,
PR	13-AUG-1999,	99US-01485653P,
PR	13-AUG-1999,	99US-01486864P,
PR	16-AUG-1999,	99US-0149368P,
PR	17-AUG-1999,	99US-01491755P,
PR	18-AUG-1999,	99US-0149428P,
PR	20-AUG-1999,	99US-0149722P,
PR	20-AUG-1999,	99US-0149723P,
PR	20-AUG-1999,	99US-0149929P,
PR	23-AUG-1999,	99US-0149302P,
PR	23-AUG-1999,	99US-0149301P,
PR	25-AUG-1999,	99US-0150566P,
PR	26-AUG-1999,	99US-0150884P,
PR	27-AUG-1999,	99US-0151065P,
PR	27-AUG-1999,	99US-0151066P,
PR	27-AUG-1999,	99US-0151080P,
PR	30-AUG-1999,	99US-0151303P,
PR	31-AUG-1999,	99US-0151348P,
PR	01-SEP-1999,	99US-01515930P,
PR	07-SEP-1999,	99US-0152363P,
PR	10-SEP-1999,	99US-0153070P,
PR	13-SEP-1999,	99US-0153758P,
PR	15-SEP-1999,	99US-0154018P,
PR	16-SEP-1999,	99US-0154039P,
PR	20-SEP-1999,	99US-01545779P,
PR	22-SEP-1999,	99US-0155133P,
PR	23-SEP-1999,	99US-0155486P,
PR	24-SEP-1999,	99US-0155659P,
PR	28-SEP-1999,	99US-0156458P,
PR	29-SEP-1999,	99US-0156596P,
PR	04-OCT-1999,	99US-0157511P,
PR	06-OCT-1999,	99US-0157753P,
PR	06-OCT-1999,	99US-0157865P,
PR	07-OCT-1999,	99US-0158029P,
PR	08-OCT-1999,	99US-01582032P,
PR	12-OCT-1999,	99US-0158369P,
PR	13-OCT-1999,	99US-0159293P,
PR	13-OCT-1999,	99US-0159294P,
PR	13-OCT-1999,	99US-0159295P,
PR	14-OCT-1999,	99US-0159329P,
PR	14-OCT-1999,	99US-0159330P,
PR	14-OCT-1999,	99US-0159331P,
PR	14-OCT-1999,	99US-0159637P,
PR	18-OCT-1999,	99US-0159638P,
PR	18-OCT-1999,	99US-0159584P,
PR	21-OCT-1999,	99US-0160741P,
PR	21-OCT-1999,	99US-0160767P,
PR	21-OCT-1999,	99US-0160768P,
PR	21-OCT-1999,	99US-0160770P,
PR	21-OCT-1999,	99US-0160814P,
PR	21-OCT-1999,	99US-0160818P,
PR	22-OCT-1999,	99US-0160980P,
PR	22-OCT-1999,	99US-0160981P,
PR	22-OCT-1999,	99US-0160989P,
PR	25-OCT-1999,	99US-0161404P,
PR	25-OCT-1999,	99US-0161405P,
PR	25-OCT-1999,	99US-0161440P,
PR	26-OCT-1999,	99US-0161359P,
PR	26-OCT-1999,	99US-0161360P,
PR	26-OCT-1999,	99US-0161361P,
PR	28-OCT-1999,	99US-0161362P,
PR	28-OCT-1999,	99US-0161992P,
PR	28-OCT-1999,	99US-0161993P,
PR	29-OCT-1999,	99US-0162114P,

[illegible]

RESULT 26

RESULT 26	
ABU48997	
ID	ABU48997 standard; protein; 475 AA.
XX	
XX	
AC	ABU48997;
XX	
DT	19-JUN-2003 (first entry)
XX	
DE	Protein encoded by Prokaryotic essential gene #34424.
XX	
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS	Ureaplasma urealyticum.
XX	
EN	WC200277183-A2.
PD	03-OCT-2002.
XX	
PF	21-MAR-2002; 2002WO-US009107.
XX	
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072851.
PR	06-MAR-2002; 2002US-0362699P.
PA	(ELIT-) ELITRA PHARM INC.
XX	
PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI	Wall D, Trawack JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	
DR	WPI; 2003-029926/02..
DR	N-PFSDB, ACAS2767.
XX	
PT	New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.


```
Db      351 GAKVLCGGERDESEKYLGGYVQETFTDTPDMKIMKEEIFFGVLPFIKFDLDEATE 410
QY      397 IANKSDPGLQSGVETRD-FOKAFDIANKLEVTGTHNNKXGRGP-NFPFLGLK-GSGAG 453
Db      411 LANDTEYGLAAYVTDKILARAFRAKALEAGIWNDCVHAAPDLPFGGVQSSGIG 470
QY      454 VQ-GIRYSIEAMTNVKSIVL 472
Db      471 REHGKYGLEYTEIKTVI 490

RESULT 28
AAG40914
ID AAG40914 standard; protein; 293 AA.
XX
AC AAG40914;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 50830.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-01334256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147600P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
```

PR	09-AUG-1999;	99US-0147493P.	QY	189	LAKAFAGAGLPGAGVNTITGRSGEIGDYIVHEBEVNFINTGTPVQRIKGLAGRRPIM	248
PR	09-AUG-1999;	99US-0147933P.	Db	1	MVCHLAGPFGKGLISCVTIGKSGEIGDPLFMHPGVNCSFTGG-DTGIALSKAGNVPLQ	59
PR	10-AUG-1999;	99US-0148171P.	QY	249	LBHGGKNGIVLADADLADLNAKQIVAGAYDYSQORTAKRVLVVEVDELAELKISNV	308
PR	11-AUG-1999;	99US-0148319P.	Db	60	MELGGKACIVLEADLADLVSANIVKGFSGYSQORTAKRVLLINESIADAVVQVNAKL	119
PR	12-AUG-1999;	99US-0148341P.	QY	309	AKLSVGDPDFNATVTPVIDNSADPFISLVVDAROKGAKELNEFKRGRLLTFLGLFDHVT	368
PR	13-AUG-1999;	99US-0148565P.	Db	120	AKLVGPPEDSDITPVVTESANFIEGLVNDAKEGATFCQYRREGNLIWFLLLDHR	179
PR	16-AUG-1999;	99US-0148684P.	QY	369	LDMLAMEEPFGPILPIIRKVDABEAVAIANKSDSFGLOSSVFTDFOKAFDIANKLEVGT	428
PR	17-AUG-1999;	99US-0149175P.	Db	180	PMRIAMEEPFGPVLPIRINSVEEGIHHCNANFGLQCIPTFDINKALLISDAMETGT	239
PR	18-AUG-1999;	99US-0149426P.	QY	429	VHNNKTRGPDNPFPLGKSGAGVGIGIYSIEAMTVKSVILDM	474
PR	20-AUG-1999;	99US-0149722P.	Db	240	VQINSAPARGEDHFPFQGLKSGIGSQGITNSINMMTKVKSTVINL	285
PR	20-AUG-1999;	99US-0149929P.	RESULT 29			
PR	23-AUG-1999;	99US-0149902P.	AAE39889			
PR	23-AUG-1999;	99US-0149930P.	ID	AAE39889	standard; protein; 493 AA.	
PR	25-AUG-1999;	99US-0150566P.	XX	AC	AAE39889;	
PR	26-AUG-1999;	99US-0150884P.	XX	AC	AAE39889;	
PR	27-AUG-1999;	99US-0151065P.	DT	18-DEC-2003	(first entry)	
PR	27-AUG-1999;	99US-0151066P.	DE	Human	aldehyde dehydrogenase domain consensus peptide.	
PR	30-AUG-1999;	99US-0151303P.	XX	Human	cytochrome p450; dehydrogenase/reductase; osteomalacia; psoriasis;	
PR	31-AUG-1999;	99US-0151438P.	XX	Human	lipoxigenase; hydratease; proliferative disorder; haematopoietic disorder;	
PR	01-SEP-1999;	99US-0151930P.	XX	Human	differentiative disorder; carcinoma; sarcoma; leukaemia; immune disorder;	
PR	07-SEP-1999;	99US-0152363P.	XX	Human	anorexia nervosa; neural degeneration; muscular disorder; lipid disorder;	
PR	10-SEP-1999;	99US-0153070P.	XX	Human	multiple sclerosis; encephalomyelitis; metabolic disorder; osteoporosis;	
PR	13-SEP-1999;	99US-0153758P.	XX	Human	myasthenia gravis; bone metabolism; immunomodulator; obesity; anorectic;	
PR	15-SEP-1999;	99US-0154018P.	XX	Human	eating disorder; osteodystrophy; arthritis; diabetes; diabetes; rickets;	
PR	16-SEP-1999;	99US-0154039P.	XX	Human	milk fever; diabetes mellitus.	
PR	20-SEP-1999;	99US-0154779P.	OS	Homo sapiens.		
PR	22-SEP-1999;	99US-0155139P.	XX	US2003032658-Al.		
PR	23-SEP-1999;	99US-0155486P.	XX	15-MAY-2003.		
PR	24-SEP-1999;	99US-0155653P.	XX	20-JUN-2002; 2002US-00175696.		
PR	28-SEP-1999;	99US-0156458P.	XX	02-FEB-2001; 2001US-0266140P.		
PR	29-SEP-1999;	99US-0156596P.	XX	04-FEB-2002; 2002US-00067668.		
PR	04-OCT-1999;	99US-0157117P.	XX	(MEYE/) MEYERS R E.		
PR	06-OCT-1999;	99US-0157753P.	XX	(GLUC/) GLUCKSMANN M A.		
PR	07-OCT-1999;	99US-0158023P.	XX	(RUDO/) RUDOLPH-OWEN L A.		
PR	08-OCT-1999;	99US-0158232P.	XX	Meyers RE, Glucksmann MA, Rudolph-Owen LA;		
PR	12-OCT-1999;	99US-0158369P.	XX	WPI; 2003-765490/72.		
PR	13-OCT-1999;	99US-0159293P.	XX	New cytochrome P450 family members, dehydrogenase/reductase protein,		
PR	13-OCT-1999;	99US-0159294P.	XX	lipoxigenase family member and human hydratase useful for treating		
PR	14-OCT-1999;	99US-0159329P.	XX	cellular proliferative and/or differentiative disorders.		
PR	14-OCT-1999;	99US-0159330P.	XX	Disclosure; Fig 16; Opp; English.		
PR	14-OCT-1999;	99US-0159331P.	XX	The invention relates to isolated cytochrome p450 family members such as		
PR	14-OCT-1999;	99US-0159637P.	XX	33312, 33302 and 32579, dehydrogenase/reductase protein such as 21509,		
PR	18-OCT-1999;	99US-0159584P.	XX	33770, lipoxigenase family member such as 46638 and human hydratase such		
PR	21-OCT-1999;	99US-0160741P.	XX	as 50090. The invention is useful as diagnostic targets and agents for		
PR	21-OCT-1999;	99US-0160767P.	XX	controlling one or more of cellular proliferative and/or differentiative		
PR	21-OCT-1999;	99US-0160770P.	XX	disorders such as carcinoma, sarcoma or haematopoietic disorders such as		
PR	21-OCT-1999;	99US-0160814P.	XX	leukaemia, immune disorders such as diabetes mellitus, arthritis,		
PR	21-OCT-1999;	99US-0160815P.	XX	multiple sclerosis, encephalomyelitis, myasthenia gravis, psoriasis, etc.		
PR	22-OCT-1999;	99US-0160980P.	XX			
PR	22-OCT-1999;	99US-0160981P.	XX			
PR	22-OCT-1999;	99US-0160989P.	XX			
PR	25-OCT-1999;	99US-0161404P.	XX			
PR	25-OCT-1999;	99US-0161405P.	XX			
PR	25-OCT-1999;	99US-0161406P.	XX			
PR	26-OCT-1999;	99US-0161359P.	XX			
PR	26-OCT-1999;	99US-0161360P.	XX			
PR	26-OCT-1999;	99US-0161361P.	XX			
PR	28-OCT-1999;	99US-0161920P.	XX			
PR	28-OCT-1999;	99US-0161922P.	XX			
PR	28-OCT-1999;	99US-0161933P.	XX			
PR	29-OCT-1999;	99US-0162142P.	XX			

Query Match 32.8%; Score 784.5; DB 3; Length 293;
 Best Local Similarity 52.4%; Pred. No. 5.3e-57;
 Matches 150; Conservative 51; Mismatches 84; Indels 1; Gaps 1;

CC or metabolic disorders such as obesity, anorexia nervosa, cachexia, lipid
 CC disorders and diabetes. The invention is useful for controlling disorders
 CC associated with abnormal fatty acid biosynthesis and neural degeneration.
 CC The invention is useful for controlling disorders associated with defects
 CC in fatty acid oxidation or proliferation or muscular disorders and for
 CC treating disorders associated with bone metabolism such as osteoporosis,
 CC osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, milk
 CC fever. The present sequence is human aldehyde dehydrogenase domain
 XX consensus peptide
 XX
 SQ Sequence 493 AA;

Query Match 32.5%; Score 777; DB 7; Length 493;

Best Local Similarity 38.5%; Pred. No. 4.7e-56;

Matches 193; Conservative 92; Mismatches 166; Indels 50; Gaps 17;

QY 12 EWKSSV--QTEILSPIDSS--LGFVPAMTREFVDHAKAGREAL---PAWAALTYPERA 65
 DB 1 EWDSASGKTEFVNPNANKGVIGRVPETAEDVDAVKAKEAFKSGPNWAKVPASERA 60
 QY 66 QYHKAADIIEERDEEATVLAKEISK--AYNASVTEVVTADLRVAA-----EEG 115
 DB 61 RIILKLADLIERDELALEALDILCKPLAEAKGDTVEGRAIDEIRYVAGWAKLMGER 120
 QY 116 I--RLSTADSGKMDASTGHKLVIRQPVGIVLAIPYVYNLSSGIAPALIGNV 173
 DB 121 VIPSLATDGD-----ELNVTREPLGVGVISPNPPLLALWKLAPALAAAGT 170
 QY 174 VMEKPTQGSVGLVAKAFAG--LPAGVENTITGRSGIGDIYVEHEEVNFINTG 230
 DB 171 VLIKPEOTPLTALLAELEBAGANNLPKGVNVVPGFAGVGGALLSHPDIDKLSFTG 230
 QY 231 STPVSGRIGKLA--GMRPTMLEGGKDGAGIVLADADLNAKQIVAGAYDYGQRTAI 287
 DB 231 STEVGKLIWEAAAKNLKVTLELGGKSPVIVFDADLDKAVRIVFAGNAGQVCIAP 290
 QY 288 KRVLVVEVDELAEKTSENVAKLS--VGDPDNAT--VTPVIDNSAD--FIESLVVDARQ 343
 DB 291 SRELVHESITDEFVEKLERVKKLLIGDPLDSDTIYGLFISEQQFDRVLWSYEDGKE 350
 QY 344 KGNK-----ELNEFKRDGRLLTPGLFDHVLDMKLAWEEPCTPLPIIRVKDAEAV 395
 DB 351 EGAKVLCCGEDESEKYLGGYVQFTIFTDTPDKMKKEEIFGVLPIIKFDLDSAI 410
 QY 396 AIANKSDGLQSSVFTRD--FOKAFDIANKLEVGTVMHNKTKGRGPD--NPFPLGLK--GSGA 452
 DB 411 ELANDTEYGLAAYFTKDLARAFRAKALEAGIVWVNDVCVHAERPOLPFGVGKSSGI 470
 QY 453 GVO--GIRYSIEAMNKKSVIL 472
 DB 471 GREHGKYGLEETEIKVTI 491

RESULT 30

ID AAB79356 standard; protein; 484 AA.

XX

AC AAB79356;

XX

DT 30-APR-2001 (first entry)

XX

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:228.

XX

KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX

OS Corynebacterium glutamicum.

XX

PN WO200100844-A2.

XX

PD 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB000943.

XX

PR 25-JUN-1999; 99US-0141031P.

PR 08-JUL-1999; 99DE-01031412.

PR 08-JUL-1999; 99DE-01031413.

PR 08-JUL-1999; 99DE-01031419.

PR 08-JUL-1999; 99DE-01031420.

PR 08-JUL-1999; 99DE-01031424.

PR 08-JUL-1999; 99DE-01031428.

PR 08-JUL-1999; 99DE-01031431.

PR 08-JUL-1999; 99DE-01031433.

PR 08-JUL-1999; 99DE-01031434.

PR 08-JUL-1999; 99DE-01031510.

PR 08-JUL-1999; 99DE-01031562.

PR 08-JUL-1999; 99DE-01031634.

PR 09-JUL-1999; 99DE-01031680.

PR 09-JUL-1999; 99DE-01032227.

PR 09-JUL-1999; 99DE-01032230.

PR 09-JUL-1999; 99US-0143208P.

PR 14-JUL-1999; 99DE-01032973.

PR 14-JUL-1999; 99DE-01033005.

PR 27-AUG-1999; 99DE-01040765.

PR 31-AUG-1999; 99US-0151572P.

PR 03-SEP-1999; 99DE-01042076.

PR 03-SEP-1999; 99DE-01042079.

PR 03-SEP-1999; 99DE-01042086.

PR 03-SEP-1999; 99DE-01042087.

PR 03-SEP-1999; 99DE-01042088.

PR 03-SEP-1999; 99DE-01042095.

PR 03-SEP-1999; 99DE-01042123.

PR 03-SEP-1999; 99DE-01042125.

XX (BADI) BASF AG.

XX

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-061975/07.

XX N-PSDB; AAF71473.

XX

PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes.

XX

PS Claim 20; Page 464-466; 1246pp; English.

XX

CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243
 CC to AAB 79633 which are involved in carbon metabolism and energy
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for
 CC expression in host cells and production or modulation of production of
 CC fine chemicals, such as, an organic acid, a proteinogenic or
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a
 CC polyketide, or an enzyme. The presence of (I) or SMP proteins (III)
 CC encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to C.
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in
 CC evolutionary studies, in determining SMP protein regions required for
 CC function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH)

XX Sequence 484 AA;

SQ

Query Match 31.0%; Score 741.5; DB 4; Length 484;

Best Local Similarity 36.9%; Pred. No. 4.2e-53;
Matches 174; Conservative 93; Mismatches 184; Indels 21; Gaps 6;

QY 6 QNVNGEW-----KSSVNOIEILSPDSSSLGFPVPMTRREVDHAKAGREALPAWAL 59
DB 12 KHLIGGWGVEGNSDRISTN---INPYDSSVTAESKQASIAADVDAEYAKQAQAEWAAT 67
QY 60 TVYERAQYLHKAADIIRDKKEIATVLAKESIKAYNASVTEVVTADLIRYAAEGIRLS 119
DB 68 PAERSAIIYRAAELEHREIEVWLIKESGSTRSKANLEITLIGNITKESASFGRV- 126
QY 120 TSADEGKMDASTGHKLAIVRRQPVGLVLAIPYVNVNLSGSKIAPALIGNVVMFKPP 179
DB 127 -----HGRISPSNTPGKENRVYRVAKVGVVISPWFPNLSIRSVAPALAVNAVVIKPA 182
QY 180 TQGSVSLVL-AKAPAEAGLPAGVFNITITGRSEIGDYIVEHEEVNFNFTGSTPVGQRI 238
DB 183 SDTPVTGVPVIRPEAGVPAGVISTVAGAGSEIGDFHVAVPKLIISFTGSTPVGRRV 242
QY 239 GKLA-----GMRPTMLELGGKAGIVLADADLNAAKQIVAGDYDQGRCTAIRVLWE 294
DB 243 GELAINGGPMKTVALELGNAPFVVLADADIDAAQAAVAGFHLQGLQICMSINRVIDA 302
QY 295 EVADELAEKISENVAKLSVGDPPDNAT-VTPVIDNDSADFTESLVVDARQKAKELNEFK 353
DB 303 AVHDEFLEKFEVAVKNIPTGDSAEGLVGPVINDSGLKELKELAKKAGATVQVEGP 362
QY 354 RDGRLTPGLFDHVTLDKLAWEPPGPIPIIRVKDAEEAIAANKSDFGLOQSVFTRD 413
DB 363 IEGRLVHPVFSVTSDEIAREEIFGLPLISVLKADDAHAELANASDFGLSAVWSKD 422
QY 414 FQKAFDIANKLEVGVTHNNKTGRGPNPFPLGLKSGAGVQGIIRYSIBAMT 465
DB 423 IDRAAQFALQIDSGMVHINDLTVDNDEPHVMFGSKNSGLGRFNGDWAIEEFT 474

RESULT 31
AAG2676
ID AAG2676 standard; protein; 484 AA.
AC AAG2676;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6430.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
FN EF1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
XX
PR 07-APR-2000; 2000JP-00159162.
XX
PR 03-AUG-2000; 2000JP-00280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
FI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
DR N-PSDB; AAH67895.
XX
ET Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX

PS Claim 17; SEQ ID NO 6430; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC Coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described in the
CC exemplification of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the European Patent Office
XX
SQ Sequence 484 AA;

Query Match 31.0%; Score 741.5; DB 4; Length 484;
Best Local Similarity 36.9%; Pred. No. 4.2e-53;
Matches 174; Conservative 93; Mismatches 184; Indels 21; Gaps 6;

QY 6 QNVNGEW-----KSSVNOIEILSPDSSSLGFPVPMTRREVDHAKAGREALPAWAL 59
DB 12 KHLIGGWGVEGNSDRISTN---INPYDSSVTAESKQASIAADVDAEYAKQAQAEWAAT 67
QY 60 TVYERAQYLHKAADIIRDKKEIATVLAKESIKAYNASVTEVVTADLIRYAAEGIRLS 119
DB 68 PAERSAIIYRAAELEHREIEVWLIKESGSTRSKANLEITLIGNITKESASFGRV- 126
QY 120 TSADEGKMDASTGHKLAIVRRQPVGLVLAIPYVNVNLSGSKIAPALIGNVVMFKPP 179
DB 127 -----HGRISPSNTPGKENRVYRVAKVGVVISPWFPNLSIRSVAPALAVNAVVIKPA 182
QY 180 TQGSVSLVL-AKAPAEAGLPAGVFNITITGRSEIGDYIVEHEEVNFNFTGSTPVGQRI 238
DB 183 SDTPVTGVPVIRPEAGVPAGVISTVAGAGSEIGDFHVAVPKLIISFTGSTPVGRRV 242
QY 239 GKLA-----GMRPTMLELGGKAGIVLADADLNAAKQIVAGDYDQGRCTAIRVLWE 294
DB 243 GELAINGGPMKTVALELGNAPFVVLADADIDAAQAAVAGFHLQGLQICMSINRVIDA 302
QY 295 EVADELAEKISENVAKLSVGDPPDNAT-VTPVIDNDSADFTESLVVDARQKAKELNEFK 353
DB 303 AVHDEFLEKFEVAVKNIPTGDSAEGLVGPVINDSGLKELKELAKKAGATVQVEGP 362
QY 354 RDGRLTPGLFDHVTLDKLAWEPPGPIPIIRVKDAEEAIAANKSDFGLOQSVFTRD 413
DB 363 IEGRLVHPVFSVTSDEIAREEIFGLPLISVLKADDAHAELANASDFGLSAVWSKD 422
QY 414 FQKAFDIANKLEVGVTHNNKTGRGPNPFPLGLKSGAGVQGIIRYSIBAMT 465
DB 423 IDRAAQFALQIDSGMVHINDLTVDNDEPHVMFGSKNSGLGRFNGDWAIEEFT 474

RESULT 32
AAU34671
ID AAU34671 standard; protein; 482 AA.
XX
AC AAU34671;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #252.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Escherichia coli.
XX
FN WO200170955-A2.
XX
PD 27-SEP-2001.
XX

PF 21-MAR-2001; 2001WO-US009180.
 XX
 XX 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS52530.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Example 3; SEQ ID NO 10264; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 482 AA;
 SQ
 Query Match 30.3%; Score 726; DB 4; Length 482;
 Best Local Similarity 34.3%; Pred. No. 8.2e-52;
 Matches 164; Conservative 92; Mismatches 206; Indels 16; Gaps 7;
 QY 6 QNVVNGKSSV--QIEILSPIDSSLSGFPVPMTEEVVDHAKGREALPAWALTYVE 63
 DB 12 QALINGELDANNGEALDVTNPANGDKLGSVPKMGADETRAAIDANRALPAWALTAKE 71
 QY 64 RAQYLKKAADIIEKDEIATVLAKEISKAYNASVEVVRVADLIRYAAEEGRLSTSD 123
 DB 72 RATILRNWFLNMEHQDDIARLTLEQKPLAEKGEISYASIEFWAEGRKRYGTI 131
 QY 124 EGKMDASTGHKLARIPGVGLIATAPYNPVNLSSGIAPALIGNVNWFKPTQGS 183
 DB 132 PGHQAD-----KRLIVIKPIQVTAATIPWNPFAAMITKAGPALAGCTWLKASQTP 186
 QY 184 VSGVLAKFAEAGLPAGVNTITRSGEIGDIVVHEEVNFINFTGSTPVGRIQKLAG 243
 DB 187 FSALALAEIATRAGVPAGVNFVVTGSAGAVNGLTSNPLVRKLSFTGSTIGRLMEQCA 246
 QY 244 --MRPMLLEGGKADGVLADADLNAKQIVAGADYSGQRCTAIRVLVVEVADELA 301
 DB 247 KOIKVSVLELGNAPFIVFDADLDKAVEGALASKFRNAGQTCVCAARLVYQGVYDRFA 306
 QY 302 EKISNVAKLSVGDPFDN-ATVTPVIDNSADFSIESLVVDARQKGAEL---NEFKRDR 357
 DB 307 EKLQAVSKLHIGDLDNGVTIGPLIDEKAVANVEEHIADELEKGVVCGKAHERGN 366
 QY 358 LLTPGLFDHVTLDKLAWESEPGPILPIIRVDAEVAIAANKSDFGLQSSVFTDRFOKA 417

Db 367 PFQPTILVDVPANAKVSKETFGFLAPLRFKDEADVIAQNOTFGLAAYFADLSRV 426
 QY 418 FPIANKLEVGTTHINKTG-RGPDNPFPLGLKSGAGVQGIYSIBAMNVKSIVLDM 474
 Db 427 FRVGEALEYGVGIN--TGIISNEVAFPGGIRKASGLRGESKYGIEDYLEIKYMCIGL 482
 RESULT 33
 ABU28726
 ID ABU28726 standard; protein; 482 AA.
 XX
 AC ABU28726;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 XX Protein encoded by Prokaryotic essential gene #14253.
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Escherichia coli.
 OS WO200277183-A2.
 FN 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA32596.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 5650; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids

Db 311 KAVGQLKVGDFGDTTGVGLIDENAVSKVQEHIEDAIQKGTLYGGQKVAELDGHFMQ 370

QY 361 P---GLFDHVTLDMLAWEPGPIPLIRVKDAEBAVAIAKSKDFGLQSSVTRDFOKA 417

Db 371 PVIIGLANDTMLCN---BETGPAVPVAKFKVEVIERANHPYGLAAIYFTKDISQA 427

QY 418 FTIANKLEVGTVHNNKTKGPD--NPFPLGKSGAGVQGGIRYSIEAMTNVKSIVLDM 474

Db 428 FQISEALBYGIIGNDGL---PSVAQAPFGGFKESGIGREGGHGFIIEYLEIKYISLGL 483

RESULT 35

AA055135

ID AA055135 standard; protein; 488 AA.

XX

AC AA055135;

XX

DT 16-OCT-2003 (revised)

DT 18-JAN-1995 (first entry)

XX

DE Heat resistant aldehyde dehydrogenase.

XX

DE aldehyde dehydrogenase; heat resistant; alcohol production;

KW Bacillus stearothermophilus.

XX

OS Geobacillus stearothermophilus.

XX

PH Key Location/Qualifiers

FT CDS 1266..2732

FT /*tag= a

FT /product= "heat resistant aldehyde dehydrogenase"

XX

XX JP06113839-A.

PN

PD 26-APR-1994.

XX

PF 30-MAR-1992; 92JP-00074631.

XX

PR 30-MAR-1992; 92JP-00074631.

PA (MARU-) MARUKAN SU KK.

XX

XX WPI; 1994-172737/21.

DR N-PSDB; AA065587.

XX

XX Heat-resistant aldehyde dehydrogenase from Bacillus stearothermophilus - and corresp. DNA, useful for alcohol prodn.

PT

XX Claim 1; Page 8-10; 11pp; Japanese.

XX

CC This sequence shows an aldehyde dehydrogenase (ALDH) (see also AA065587), that is heat resistant from Bacillus stearothermophilus and requires no acetyl CoA. The fused protein of ALDH with an alcohol dehydrogenase derived from a transformant which contains both structural genes can be used in production of alcohol. (Updated on 16-OCT-2003 to standardise OS field)

XX

XX Sequence 488 AA;

SQ

Query Match 30.2%; Score 721.5; DB 2; Length 488;

Best Local Similarity 34.4%; Pred. No. 2e-51;

Matches 166; Conservative 104; Mismatches 192; Indels 21; Gaps 9;

QY 3 KEYQNVNGEKSSV-NQIE-ILSPIDSSL-GFVPAMTRFVDHAKAGREALPAWAAAL 59

Db 8 KYTFYNGNVSSVSNVPEINPANRHIVGVVORSTLEDVNEAVTANEAQISWKR 67

QY 60 TVYERAYLHKADIIERDKETATVLAKESIKAYNASVTEVTRADLIRYAAERGIS 119

Db 68 SGVERGEYLYKAAHILEQCLQDIATMTBWMGKTLAEAKAETMRGVHILRYVAGEGAR-- 125

QY 120 TSADGGKMDASTGKLAIVRQPGVILAIAPYPNVNLSSKIALIGGNVWFKPP 179

Db 126 ---KIGDVIPESSDSEGLFTTRVPLGVGVISPNFFVPAIPKMAPALVXGNTVVLKEA 182

QY 180 TQSSVSLVLAKAEAEAGLPGAVFNITIGRSEIDGYDIVEHEEVNFINTGTPVQORIG 239

Db 193 SEYATAAKVIECHFEAGFRGVNVNVCSSVVGQGIANHFDIDIGVTTFTGNTFVGKQVG 242

QY 240 KLAGMR--PIMLELGGKADAGIVLADADLNNAAKQIVAGAYDYSGQRCTAIKRVLVVEVA 297

Db 243 RAAAFERGAQVLEMGKXNPVIVAKADLDLAVEGTISGGLRSTGQKCTATSRVFEREVY 302

QY 298 DELAEKISENAKSLVGDPPDNAT-VTPVIDNDSADFISSLVVDARQKAKELNEFR-- 354

Db 303 EPPKAKLLERVKQLKIGNGLDAETWMPFCASESFHTVLSYIEKKGSEGAkliYGGNRCL 362

QY 355 -----DGRLLTPGLFDHVTLDMLAWEPGPIPLIRVKDAEBAVAIAKSKDFGLQSSV 409

Db 363 EGELANGPFVEFTTFEDVDLQMTIAREIFQFVLIQVDSIEEAIKLANDTEYGLSASI 422

QY 410 FTRDFQKAFDIANKLEVTGVHNNKTKGPDNPFPLGLK--GSGAGVQGGIRYSIEAMTNV 467

Db 423 YTKNIGNALEFTKIDIEAGLIKVNNAETAGVEFQAPFGGMKQSSSHSREG-QAAIEFTTSI 481

QY 468 KSI 470

Db 482 KTV 484

RESULT 36

AB047807

ID AB047807 standard; protein; 488 AA.

XX

AC AB047807;

XX

DT 05-FEB-2002 (first entry)

XX

DE Listeria monocytogenes protein #511.

XX

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease.

XX

OS Listeria monocytogenes.

XX

XX WO200177335-A2.

XX

XX 18-OCT-2001.

XX

XX 11-APR-2001; 2001WO-FR001118.

XX

XX 11-APR-2000; 2000FR-00004629.

XX

XX (INSP) INST PASTEUR.

XX

XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;

PI Dussurget O, Cherouani F, Medjari H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;

XX

XX WPI; 2002-010914/01.

DR

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

PT and prevention of Listeria and related bacterial infections, and related

PT polypeptides.

XX

PS Claim 6; SEQ ID NO 512; 192pp; French.

XX

CC The present invention relates to the genome sequence of Listeria

CC monocytogenes EGD-e (see AB03041). The genome sequence and fragments of

CC it are useful for selecting probes and primers for detecting genes in L.

CC monocytogenes and related organisms, and for studying genetic

CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 488 AA;
 Query Match 30.0%; Score 718; DB 5; Length 488;
 Best Local Similarity 33.8%; Pred. No. 3.9e-51;
 Matches 161; Conservative 96; Mismatches 201; Indels 18; Gaps 6;
 QY 8 YVNGEKSSVNO--IEILSPIDSSLGFPVPMTRREVDHAKAGREALPAWALTVERA 65
 DB 16 FINGKWTGDNKETKDINPANGDVIAKTAQAGPSETKKAIAKADAFDPAWAKMELADRV 75
 QY 66 QYLHKAADIIRDEKEEIAITVLAKEISKAYNASVTVETVDTADLIRYAEEGIRLSTSADEG 125
 DB 76 KLLHKAIDLMEEKADTLAKIMTLEQCKPLKESKGEVLTGVENFRAAEARRL-----YG 130
 QY 126 GKMDASTGHKLAIVIRQPVGLVLAIPYVNVNLSGSKIAPIALIGNVVMEKPTQGSVS 185
 DB 131 ETIPAPNNHAF-IVKKQPIGVVAALTWNFPDGMVTRKLPALATGNTIVLKPSTDPLS 189
 QY 186 GLVLAFAFAGLAPAGVNTITGRSGEIGDIYVEHEEVNFTGTSPVQRIKGLAG-- 243
 DB 190 ALAIFEFREAGLPGVANIWMGSSKEIGETLTDSDVRKLTGTSTKVGQTLFKQADT 249
 QY 244 MRPTMLEGKQAGIVLADLDNAKQIVAGAYDSQRCRTAKKVLVVEVDELAEX 303
 DB 250 LKKISLELGHAFPIVFDANLDAVNDLVAAKFERNQVCVSPNFIKAKEIKETKA 309
 QY 304 ISENVAKLSVGGPFDNATVTPVIDNSADFIESLVDAKQKALENEFR-----DG 356
 DB 310 LVAKVEQLKVGKGLDNNVGLIREADIDKQKAKATEKAKVLTGGRLTGSYDKG 369
 QY 357 RLLTFGLDHVLDKMLAEEPPGPIILPIIRVKDAEEAVAIANKSDFGLQSSVTRDFQK 416
 DB 370 NFYKPTVLDNVTRKMDIFEETFGFVPLITFETEDAEIEMANDSEFGLASVFFYTKDLAR 429
 QY 417 AFDIANKLEVTGHNNKTRGPDNPFPLGLKSGCAGVGIRYSIEMNTVKSIVL 472
 DB 430 VEKVGAALEYGMVGAIEIAISNPET-PFGGVKHSFGRENGHYGMEEYIQVKFINL 484

RESULT 37
 ABU32914
 ID ABU32914 standard; protein; 488 AA.
 XX AC ABU32914;
 XX DT 19-JUN-2003 (first entry)
 XX DE Protein encoded by Prokaryotic essential gene #18441.
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX OS Listeria monocytogenes.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA36784.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 60838; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than S. aureus, S. typhimurium,
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 488 AA;
 Query Match 30.0%; Score 718; DB 6; Length 488;
 Best Local Similarity 33.8%; Pred. No. 3.9e-51;
 Matches 161; Conservative 96; Mismatches 201; Indels 18; Gaps 6;
 QY 8 YVNGEKSSVNO--IEILSPIDSSLGFPVPMTRREVDHAKAGREALPAWALTVERA 65
 DB 16 FINGKWTGDNKETKDINPANGDVIAKTAQAGPSETKKAIAKADAFDPAWAKMELADRV 75
 QY 66 QYLHKAADIIRDEKEEIAITVLAKEISKAYNASVTVETVDTADLIRYAEEGIRLSTSADEG 125
 DB 76 KLLHKAIDLMEEKADTLAKIMTLEQCKPLKESKGEVLTGVENFRAAEARRL-----YG 130
 QY 126 GKMDASTGHKLAIVIRQPVGLVLAIPYVNVNLSGSKIAPIALIGNVVMEKPTQGSVS 185
 DB 131 ETIPAPNNHAF-IVKKQPIGVVAALTWNFPDGMVTRKLPALATGNTIVLKPSTDPLS 189
 QY 186 GLVLAFAFAGLAPAGVNTITGRSGEIGDIYVEHEEVNFTGTSPVQRIKGLAG-- 243

Db 190 ALAIFEEAGLEKPGVANIWMGSSKEITGELTDSVDKRLTFTGSTKVGQTLFKQSADT 249
 QY 244 MRPIMLELGGKADGIVLADADLNAAKQIVAGADYSGQRTAJKRVLVVEEVADELAK 303
 Db 250 LKISILELGGHAPFIVFDDANLDAADVAAKERNNGQVSPNRFVAKKEKFTKA 309
 QY 304 ISENVAKLSVGDPPDNATVTPVIDNSADFTIESLVDARQKGAKEINEFKR-----DG 356
 Db 310 LVAKVEQLKVGNGLDVNVGFLIREADDKDKLNATEKGAIVLTGGGLGSDYDKG 369
 QY 357 RLITPGLFHVTLDMKLAWEPPGPILPIIRVKDAEBAVAIAKNSDFGLSSVFTDFOK 416
 Db 370 NFYKPTVLNTRKWDIFYETEGPVIPLITFETEDAEIEMANDSEFGLASYFTYKDLAR 429
 QY 417 AFDIANKLEVGTVHINKTKGRGDNFPFLGLKSGAGVQGIKRYSTRMTNVKSVIL 472
 Db 430 VEKVGAALEYGMVGAINEIASNPT-PFGGVKHSFGFNGHYGMEIYQVKFINL 484

RESULT 38

ABU31620
 ID ABU31620 standard; protein; 482 AA.

XX AC ABU31620;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #17147.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Klebsiella pneumoniae.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA35490.

XX PS Claim 25; SEQ ID NO 59544; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
 the target prokaryotic essential genes. Note: The sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 482 AA;

Query Match 29.7%; Score 711; DB 6; Length 482;
 Best Local Similarity 34.1%; Pred. No. 1.5e-50;
 Matches 163; Conservative 90; Mismatches 209; Indels 16; Gaps 7;

QY 6 QNYVNGEWKSSVN--QIBILSPIDSSLGTFVPMATREEDVHAMKAGREALPAWAAALTVE 63
 Db 12 QALIDGQWRDAPNGDVIATVNPANGQLGSPKMGADTREATAANRALPAWRALTAK 71
 QY 64 RAQVILHVAADIETDKESIAIVLAKESIKAYNASVTEVVTADLIRYABEGIRLSTAD 123
 Db 72 RANTLRWFDLMMNQDDILAMTLEQCKPLAEAKGSIYAAAFIEWFAEBGKRIYDGTI 131
 QY 124 EGGKMDASTGHKLAVIRROPVGIIVLAIAPYVPVNLGSKIAPALIGNVNVMFKPTQGS 183
 Db 132 PCHQAD-----KLLIVIKQIPGIVTAITPWPFPAMITRKAGPALAAGCTVWLFASQTP 186
 QY 184 VSGILVAKAFABAGLPAGVNTITGRSGSHIGDIYIVHEHEVNFINFTGSPVQRIKLAG 243
 Db 187 FSLALAEALANRAGIPAGVFNVTGAGAVGGLTSLNPLVKLSPTGSTEIGRQIMEQCA 246
 QY 244 --MRPIMLELGGKADGIVLADADLNAAKQIVAGADYSGQRTAJKRVLVVEEVADELA 301
 Db 247 KDIEKVSLELGGNAFFIVDDADLDKXAVEGALASKFRNAGQTCVCANRLYQDGVYDRFA 306
 QY 302 EKISENVAKLSVGDPE-DNATVTPVIDDSDAFIESLVDARQKGAKEINEFK--RDGR 357
 Db 307 EKLOQAVEKLRIQDGLQDQVTTGFLIDEKAVAKVEEHIADIAKGAKVVTGKPHALGNN 366
 QY 358 LLTFLGFLDHRVTLDMKLAWEPPGPILPIIRVKDAEBAVAIAKNSDFGLSSVFTDFOKA 417
 Db 367 FFQPTILVNVPDSAKVAKEETFGPLAPLFRFKDEADVIAQANDTEFLAAVFAFYARDLSRV 426
 QY 418 FDIANKLEVGTVHINKTKG-RGDNFPFLGLKSGAGVQGIKRYSTRMTNVKSVILDM 474
 Db 427 FRVGEALLEYGIIGIN--TGIISTEIVAPFGGVKASGLRGSKYGIEDYLEIKYMCIGL 482

RESULT 39

ABB08382

ID ABB08382 standard; protein; 491 AA.

XX AC ABB08382;

XX DT 29-AUG-2003 (revised)

XX DT 18-JUN-2002 (first entry)

XX DE PcuC amino acid sequence.

XX KW PHBA; para-Hydrobenzoate; liquid crystal polymer; LCP;

XX KW toluene monooxygenase; TMO; pcu gene; p-cresol.

XX KW Pseudomonas mendocina; KR-1.

XX XX

Db 68 TAPARARVLEAVKIDFERREIIDWIRESGS-----TRIAQIEWGAARATL 118
 QY 119 STSDEG---GKMDAST-GHKLAVIRROPVIGVILAIAPVNPVNLGSKIAPALIGNVV 174
 Db 119 ESASLPNRVHGIASINIGKESRYRAPLGVIGVISPWNPHLHARSAPALAGNAV 178
 QY 175 MKPPTQGSV-GVLAKAFABAGLPAGVFNITITGRSGEIGDYIVHEEVEVNFITGSTP 233
 Db 179 VVKPASDPTITGILLARIFEAGLPAGVLSVVVSGSAEIGDAFVEHPVALISTGSTQ 238
 QY 234 VQORTGKLAG---MRPIMELGGKXGAGVILADLDNNAKQIVAGAYDYGORTAIKR 289
 Db 239 VCENTIGRATSGEHLKHALELGGNSPFVVLADADVEQAVNAAVVGKFLHQGQICWAINR 298
 QY 290 VLVVEVEADELAEKISENVAKLSVGDPTDNATVT-PVIDNSADFIESLVVDAROKGAK 348
 Db 299 IIVEQFLEEDTRFRVERKALPYGDSKFGTVGVGVINARQLAGLKEKATAKAEGATL 358
 QY 349 LNEFKRDGLLTLGFLDHTLDMKLAMEPFGPILPIIRVKDAEEVAIAKNSDFGLQSS 408
 Db 359 LLGGEPPQGNVMPHVFNGVNTADMEIAREEIFGLVIGQSARDAEHALELANSSEYGLSSA 418
 QY 409 VTRFQKAFDIANKLEVTVHNNKTRGPDNFPFLGLKSGAGVGGIRYSIEAMTVNK 468
 Db 419 VITASLERGVQFARRIHAGMTHVNDIPVNDENAPFGKSGKSLGRFNGDMWAEIEFTTDH 478
 QY 469 SIVL 472
 Db 479 WITL 482

RESULT 41

ABU40285
 ID ABU40285 standard; protein; 480 AA.
 AC ABU40285;
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #25812.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Pseudomonas putida.
 OS WO200277183-A2.
 PN 03-OCT-2002.
 PD 21-MAR-2002; 2002WO-US009107.
 PF 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-036299P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA44155.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 68209; 1766pp; English.
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 480 AA;

Query Match 29.4%; Score 704; DB 6; Length 480;
 Best Local Similarity 34.0%; Pred. No. 5.6e-50;
 Matches 162; Conservative 91; Mismatches 209; Indels 14; Gaps 7;
 QY 6 QNYVNGEKSSVN--QIEILSPIDSSILGFPVPMTRREVDHAKMAGREALPAAULTVYE 63
 Db 12 QAVINGELDADNGQTIKVTNPATGEVIGTPKMGTAETTRAIEAADKALPAMFAUTAKE 71
 QY 64 RAQYLKHAADIIEKSEIATVLAKEISKANASVETVVTADLIRYAAEGIRLSTSD 123
 Db 72 RSALKRRWFELMIENQDGLARLMTTEQCKPLAEAKGEIAYAAASFIEWFAEAKRIYGD 131
 QY 124 EGKMDASTGHKLAVIRROPVIGVILAIAPVNPVNLGSKIAPALIGNVVMPKPTQGS 183
 Db 132 PGHQPD-----KRLIVIKQPIGVTAITPWNFPAMITKAGPALAAGCTMVLKPSQTP 186
 QY 184 VSGVLAKAFABAGLPAGVFNITITGRSGEIGDYIVHEEVEVNFITGSTPVGRIK--L 241
 Db 187 YSALALVELLAHRAGIPAGVLSVVVSGSAGEVGTGNSLVKLSFTGSTEIGRLMEECA 246
 QY 242 AGWRPIMELGGKXGAGVILADLDNNAKQIVAGAYDYGORTAIKEVLVVEVEADELA 301
 Db 247 KDIKKVSLELGGNAPFIVDDADLDKAVEGAIISKYRNNGQTCVCANRIYVQDGVYDAFA 306
 QY 302 EKISENVAKLSVGDPPFONATVT-PVIDNSADFIESLVVDAROKGAKELNEFKR-DGRL 359
 Db 307 EKLAARVAKLXINGLEBEGTTTGLIDGKAVAKVQEHIEDAVSKGKVLSCGLLEGNFF 366
 QY 360 TPGFLDHTLDMKLAMEPFGPILPIIRVKDAEEVAIAKNSDFGLQSSVTRDFQKAFD 419
 Db 367 EPTILVDVPKTAATAVAKEETFGPLAPLFRFKDEAEVIAMSNDFEGLASYFFYADMRGRFR 426
 QY 420 IANKLEVTVHNNKTR-CPDNFPLGLKSGAGVGGIRYSIEAMTVNKSVILDM 474
 Db 427 VAEALRYGMGIN--TGLISNEVAPFGIKASGLREGSKYGLIEDYLEIKYLCISV 480

RESULT 42

ABU41623
 ID ABU41623 standard; protein; 480 AA.
 XX

AC ABU41623;
 XX 19-JUN-2003 (first entry)
 XX Protein encoded by Prokaryotic essential gene #27150.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Pseudomonas syringae.
 XX WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 XX 25-OCT-2001; 2001US-0342923P.
 XX 08-FEB-2002; 2002US-00072851.
 XX 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 XX N-PSDB; ACA45493.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 69547; 1766pp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 XX the 6213 antisense sequences given in the specification where expression
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX (1) a vector comprising a promoter operably linked to the nucleic acid
 XX encoding a polypeptide whose expression is inhibited by the antisense
 XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX polypeptide or its fragment whose expression is inhibited by the
 XX antisense nucleic acid; (4) an antibody capable of specifically binding
 XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 XX proliferation or the activity of a gene in an operon required for
 XX proliferation; (7) identifying a compound that influences the activity of
 XX the gene product or that has an activity against a biological pathway
 XX required for proliferation, or that inhibits cellular proliferation; (8)
 XX identifying a gene required for cellular proliferation or the biological
 XX pathway in which a proliferation-required gene or its gene product lies
 XX or a gene on which the test compound that inhibits proliferation of an
 XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
 XX compound's activity; (11) a culture comprising strains in which the gene
 XX product is overexpressed or underexpressed; (12) determining the extent
 XX to which each of the strains is present in a culture or collection of
 XX strains; or (13) identifying the target of a compound that inhibits the
 XX proliferation of an organism. The antisense nucleic acids are useful for
 XX identifying proteins or screening for homologous nucleic acids required
 XX for cellular proliferation to isolate candidate molecules for rational
 XX drug discovery programs, or for screening homologous nucleic acids
 XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 XX the target prokaryotic essential genes. Note: The sequence data for this
 XX patent did not form part of the printed specification, but was obtained
 XX in electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 480 AA;
 XX Query Match 29.0%; Score 693; DB 6; Length 480;
 XX Best Local Similarity 33.8%; Pred. No. 4.6e-49;
 XX Matches 161; Conservative 89; Mismatches 212; Indels 14; Gaps 7;

QY 6 QNYVNGEW--KSSVNOIEILSPIDSSIGFVPAMTREVVDHMKAGREALPAWALTYVE 63
 DB 12 QAYINGQWLDADGGQSIKYNVPATNEILGTVPMGAETRAIEAADKALPAWALTAKE 71
 QY 64 RQYILHKAADIIEERKEIATVLAKE:SKAYNASVTVVVTADLIRYAAEEGRJLSTAD 123
 DB 72 RGNKLRWFELMIENQDGLLMTLEQKPLAEAKGEITYAASTFIEWFAEAKRVVGDVI 131
 QY 124 EGGKWDASTGHKLAVEROPGIVLAIPNYPNLSGSKTAPALIGNVVMEKPTQGS 183
 DB 132 PQHQPD----KGLIVIKQPIGVTAAITPNWFFAAMITRKAGPALAAGCTMVLKPSQTP 186
 QY 184 VSGVLAKAFAPAEAGLPAGVFNITITGRGSEIGDYIVHEVNFNFTGTPTVGORICK--L 241
 DB 187 FSNALALAEAEFRAGIPAGVSVVTVGSAGDLSGLTGNPVIKLSFTGTSTIGRLMAECA 246
 QY 242 AGNRPMLELGGKADGIVLADADLNAAKQIVAGAYDVSGQRCTAIKRVILVVEEVADELA 301
 DB 247 KDKKVSLELGNAPFIVFDADLDKAVEGAMISKYRNNGQTCVCANRIYVQDGYDAFA 306
 QY 302 EKISENVAKISVGDPF-DNATVTPVIDDNSADFIESLVVDARQKGAKEINEFKR-DGRLL 359
 DB 307 EKLKVAVGKUKINGLEDGITTGPLIDEKAVAKVEHIAVSGATVLTGNSLEGFF 366
 QY 360 TPGLFDRHTLDMKLAWEEPFGPIPIIRVKDAEEAVAIANKSDFGLOSSVFTDFQKAFD 419
 DB 367 EPTILVNVSKDAVAAREETFGPLAFLRFKDEAEALALANDTERGLASYFYAGNMSRVFR 426
 QY 420 IANKLEVTGTHNNKTRGPDNF-PFLGLKSGAGVQGIIRYSIEAMTVKSVILDM 474
 DB 427 VAELLEYGMVGIN--TGLISNELAPFGGIKSSGLGREGSKYGIEDYLEIKYLCLSV 480

RESULT 43
 ABU42152
 ID ABU42152 standard; protein; 480 AA.
 XX AC ABU42152;
 XX DT 19-JUN-2003 (first entry)
 XX Protein encoded by Prokaryotic essential gene #27679.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Pseudomonas syringae.
 XX WO2000277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 XX 25-OCT-2001; 2001US-0342923P.
 XX 08-FEB-2002; 2002US-00072851.
 XX 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 XX N-PSDB; ACA46022.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 70076; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 480 AA;

Query Match 29.0%; Score 693; DB 6; Length 480;

Best Local Similarity 33.8%; Pred. No. 4.6e-49;

Matches 161; Conservative 89; Mismatches 212; Indels 14; Gaps 7;

QY 6 QNYVNGEW--KSSVNVQIEILSPIDSSGLFVPTREVDHAKAGREALPAAATVYVE 63
 Db 12 QAYINQWLDADGGGSKVNNPATNEILGTVPKMGAEATRRALRADKALPAMRAUTAKE 71
 QY 64 RAQYHLKAADIIRDKKEIATVLAKESKAYNASVTEVTRADLRYAAEBGIRLTSAD 123
 Db 72 RGNKLRWFELMENQDGLLNTLQKGLAEAKGKITVAASFIEWFAEEAKEVYGDVI 131
 QY 124 EGGKMDASTCHKLAVIRQPVGLVLAIPYVFNLSGSKLAPALIGNVVMFKPTQGS 183
 Db 132 PGHQPD-----KRLIVLKQPIGVTAATIPWNPFAAMITRKAGPALAAGCTWILKPASQTP 186
 QY 184 VSGILVAKAFAGAPAGVNTTIGRSGEIGDYIVEHEVNFINTGSTPVGORICK--L 241
 Db 187 FSLALAEALAEAGIPAGVSVTSGAGDIGSEITGNPIVRKLSFTGTEIGQLMAECA 246
 QY 242 AGMRPIMLEGGKAGIVLADADLNNAAQIVAGAYDSQCQRTAKRLVVEEVADELA 301
 Db 247 KDIKVYSELGGNAPFIVFDADLDKAVEGAMI SKYRNNGQTCVCANRIYVQDGVYDAFA 306
 QY 302 EKISENVAKLSVGDPF--DNATVTPVIDNDSADFTESIVUDAKQKAKELNEFR--DGRLL 359
 Db 307 EKLKAVGKLKTNGLNGLEDGTTGLDIDEKAVAKVEHIAVDVSKGATVLTGNSLSGSPF 366
 QY 360 TPGIFDHTLDMKLAWEPPGPIPIIRVKDAEEAVALANKSPDGLQSGVTRDFOKAFD 419
 Db 367 EPTILVNSKDAVARETGPAPLRFKDEAEAILANDTEFGLASYFAQMGVERVF 426
 QY 420 IANKLVGTVTHNNKTRGPDNF--PELKGSGAGVGQIRYSIEAMTNVKSIVLDM 474
 Db 427 VAAEALYGVMGVGIN--TGLISNELAPFGGIKSGLGREGSKYGLIEDYLEIKVLCISV 480

RESULT 44

ABU47293
 ID ABU47293 standard; protein; 482 AA.
 AC ABU47293;
 XX
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #32820.
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS *Salmonella typhimurium*.
 PN W0200277183-A2.
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JB, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA51163.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 75217; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 482 AA;

Db 194 LTALAAAEALQAGVPPGALNVWGNAPETIGDALLTSPQVRKITTGTSTAVGKIMAAAA 253
QY 244 --MRPIMLEGGKDGAGIVLADLDNAAKQIVAGAYDYGORCTATKRVLVVEEVADELA 301
Db 254 PTVKVGLEGGNAPSIFDDADLDVAVKGTIAAFNRSGQTCVCAARVLVDQGYDKFA 313
QY 302 EKISENVAKLSVGDPF--DNATVTPVIDDNSADFIELSVVDARQKAKELNEEFKRDGRILT 360
Db 314 EAFSEAVQKLEVDGFRDGTQGLINDAAVQKVTFVQDAVSKGAKIIGGKRHSLGMT 373
QY 361 ---PCLFDHYTLDMKLAMEEPPGILPIIRVKDAEAVAIANKSDFGLOSSVTFDRQKA 417
Db 374 FYEPTVIRDVSDNIMSKSEEPGVPALIRPKTEEDAIRANDTIAGLAAYIFTNSVQRS 433
QY 418 FDIANKLEVGTGHNNKTRGPDNPFPLGLKSGAGVGQIRYSIEAMTVNKVIVLDM 474
Db 434 WEVFEALYGLVGNVGL-ISTEVAPEGVGKQGLGREGSKGMDEYLEIKYVCLGDM 490

RESULT 49
ID AAG23558
XX AAG23558 standard; protein; 495 AA.
AC AAG23558;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 26911.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144844P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155655P.
PR 28-SEP-1999; 99US-0156459P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.

```

PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637F.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160984P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      28.2%; Score 674.5; DB 3; Length 495;
Best Local Similarity 34.5%; Pred. No. 1.7e-47;
Matches 165; Conservative 89; Mismatches 209; Indels 15; Gaps 7;

Qy 6 QNYVNGEWSVNQ--IELSPIDDSLSGFVPANTREEDVDMKAGKAGREALPAWALTVE 63
Db 21 QELIGGKWLSDYDKTIKVNYPATGEIADVACMKETKENDAIASSYEAFTSWRLTAGE 80

Qy 64 BAQYHKAADIETKKEIATVLAKESKAYNASVTEVVTADLIRYAAEGIRLSTAD 123
Db 81 RSKVLRRWYDLIAHKEELQTLTEQCKPLKEAIGEAVAGSFIYYAEAKV----- 135

Qy 124 EGGKMDASTGKLVIRQPVGIIVLAIPYVNLSSKIAIPALIGNVVWFPPPTQGS 183
Db 136 YGDIIPPNLSDRLILVLPQVGVVGAITPWFPLAMITRKVGPALASGCTVVVKFSELTP 195

Qy 184 VSGILVLAKAFAGLPAGVENTITGRSGEIGDYIVIEHEEVNFINETGSPVQRIKLAG 243
Db 196 LTALAAELALQAGVPPGALNVVNGNAPIEIGDALLTSFQVRKITFTGTAVGKKLMAAA 255

Qy 244 --MRPIMLEIGGKAGIYVLAADLDNAAKQIVAGAYDYGQRCCTAIKRVLVVEEVADELA 301
Db 256 PTVKKSLEIGGNAPSIVFDDADLDVAVKGTLLAKFRNSGOTCVCANRVLVQDGIYDKFA 315

Qy 302 EKISENVAKLSVGDPF-DNATVTEVIDDMSADFTESLVVDARQKAKELNEFKDGRLLT 360
Db 316 EAFSEAVQKLEVGDFRGTGTQGLINDAAVQVETTFQDAVSKGAKIIIGKSHSLGMT 375

Qy 361 ---PGLFDHVLDMKLAEEFPFGPILPIIRVKDAEEAAVAIAKNSDFGLQSSVETRDFOKA 417
Db 376 FYEFTVIRVSDNMIMSKKEIFGVPAPLIRFKTEEDAIRINDTTAGLAAYITNSVQKS 435

Qy 418 FDIANKLEVGTTHNNKTGRGNPFPLGLKSGAGVQGIKRYSIEMNTNKSIVL-DW 474
Db 436 WRVFEALEYGLVGVNNEGL-ISTEVAPFGVKGQGLGREGSKYGMDEYLEIKYVCLGDM 492

```

Search completed: April 13, 2004, 14:20:58
Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:17:03 ; Search time 45 seconds
(without alignments)
3330.470 Million cell updates/sec

Title: US-09-868-195-12

Perfect score: 2393

Sequence: 1 LTKXQNTYVNGMKSSVNOI.....GIRYSIEMTNKVSIVLDMK 475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_ivirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2391	99.9	475	2 Q8RP84	Q8RP84 streptococc
2	2382	99.5	475	16 Q8E0B1	Q8E0B1 streptococc
3	2379	99.4	475	16 Q8E5Y8	Q8E5Y8 streptococc
4	1812	75.7	474	16 Q97QT0	Q97QT0 streptococc
5	1802	75.3	474	16 Q8DPF5	Q8DPF5 streptococc
6	1645	68.7	475	16 Q8P0F5	Q8P0F5 streptococc
7	1639	68.5	475	16 Q9Z67	Q9Z67 streptococc
8	1618	67.6	467	16 Q8K707	Q8K707 streptococc
9	1383	57.8	482	16 Q97D25	Q97D25 clostridium
10	1338	55.9	479	16 Q81UL6	Q81UL6 bacillus an
11	1333	55.7	479	16 Q81HE6	Q81HE6 bacillus ce
12	1323	55.3	482	16 Q8XHP4	Q8XHP4 clostridium
13	1244	52.0	481	16 Q9KA00	Q9KA00 bacillus ha
14	1171.5	49.0	496	10 Q9SNX8	Q9SNX8 apium grave
15	1168.5	48.8	496	10 Q9ZUG8	Q9ZUG8 arabidopsis
16	1145.5	47.9	507	10 Q8S4Y9	Q8S4Y9 oryza sativ

17	1138.5	47.6	508	10 Q84PC4	Q84PC4 oryza sativ
18	1047.5	43.8	478	16 Q8EVT9	Q8EVT9 mycoplasma
19	1027.5	42.9	496	10 Q8LX61	Q8LX61 triticum ae
20	1026.5	42.9	496	10 Q8LSJ9	Q8LSJ9 triticum ae
21	868.5	36.3	475	16 Q9PQC9	Q9PQC9 ureaplasma
22	786	32.8	478	17 Q8ZV31	Q8ZV31 pyrobaculum
23	752	31.4	470	17 Q97XA5	Q97XA5 sulfolobus
24	751	31.4	468	17 Q97X55	Q97X55 sulfolobus
25	750.5	31.4	486	16 Q8EMK2	Q8EMK2 oceanobacil
26	750	31.3	470	17 Q97X59	Q97X59 sulfolobus
27	742	31.0	497	16 Q8CV96	Q8CV96 oceanobacil
28	741.5	31.0	455	17 Q27059	Q27059 methanobact
29	741.5	31.0	484	16 Q8NMB0	Q8NMB0 corynebacte
30	741.5	31.0	489	16 Q9HUR4	Q9HUR4 pseudomonas
31	739	30.9	506	17 Q8U2S5	Q8U2S5 pyrococcus
32	737	30.8	476	16 Q66573	Q66573 aquifex aeo
33	736.5	30.8	491	2 Q59702	Q59702 pseudomonas
34	736	30.8	474	16 Q81QX6	Q81QX6 bacillus an
35	735	30.7	482	16 Q9KC36	Q9KC36 bacillus ha
36	734	30.7	471	17 Q8TWC7	Q8TWC7 methanopyru
37	725	30.3	499	16 Q8FUE1	Q8FUE1 corynebacte
38	721.5	30.2	483	16 Q81ZE2	Q81ZE2 bacillus an
39	720	30.1	488	16 Q92DA3	Q92DA3 listeria in
40	718	30.0	488	16 Q8Y819	Q8Y819 listeria mo
41	717	30.0	482	16 Q8X950	Q8X950 escherichia
42	716.5	29.9	493	17 Q9HK01	Q9HK01 thermoplasma
43	716	29.9	474	16 Q81DV8	Q81DV8 bacillus ce
44	713.5	29.8	480	17 Q9HQ22	Q9HQ22 halobacteri
45	713	29.8	501	1 Q57693	Q57693 thermoprote
46	711.5	29.7	472	16 Q8EMH9	Q8EMH9 oceanobacil
47	711.5	29.7	483	16 Q81IM7	Q81IM7 bacillus ce
48	711	29.7	475	16 Q8ELI8	Q8ELI8 oceanobacil
49	710.5	29.7	489	17 Q9HMJ6	Q9HMJ6 halobacteri
50	710.5	29.7	498	16 Q8YQP9	Q8YQP9 anabaena sp

ALIGNMENTS

RESULT 1
Q8RP84 ID Q8RP84 PRELIMINARY; PRT; 475 AA.
AC Q8RP84;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21843113; PubMed=11854208;
RA Hughes M.J., Moore J.C., Lane J.D., Wilson R., Pribul P.K.,
RA Younes Z.N., Dobson R.J., Everest P., Reason A.J., Redfern J.M.,
RA Greer F.M., Paxton T., Panico M., Morris H.R., Feldman R.G.,
RA Santangelo J.D.;
RA "Identification of Major Outer Surface Proteins of Streptococcus
RT agalactiae.";
RL Infect. Immun. 70:1254-1259 (2002).
CC EMBL; AF439646; AL85685.1; -
CC -I- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR Pfam; PF00171; algedh; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 475 AA; 51:98 MW; C525151C88B6AC01 CRC64;
Query Match 99.9%; Score 2391; DB 2; Length 475;

Best Local Similarity 99.8%; Pred. No. 7.9e-140;
Matches 474; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKKEYQNVNGEWSKSSVNOIEILSPIDSSISLGFVPAMTREVVDHAKAGREALPAAALT 60
Db 1 MTKKEYQNVNGEWSKSSVNOIEILSPIDSSISLGFVPAMTREVVDHAKAGREALPAAALT 60

QY 61 VYERAQYLHKAADIIERDKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120
Db 61 VYERAQYLHKAADIIERDKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120

QY 121 SADEGKMDASTGHKLAVTRPPQVIGVLAIPNYPNLSGSKIAPALIGNVVMFKPPT 180
Db 121 SADEGKMDASTGHKLAVTRPPQVIGVLAIPNYPNLSGSKIAPALIGNVVMFKPPT 180

QY 181 QGSVSLVLAKAFAPAGLPGVFNITGRGSEIGDYIVHEEVNFNFTGSTPVQRIK 240
Db 181 QGSVSLVLAKAFAPAGLPGVFNITGRGSEIGDYIVHEEVNFNFTGSTPVQRIK 240

QY 241 LAGMRPIMLELGGKDGAGIVLADADLNAKQIVAGAYDSGORCTAIKRVLVVEEVADEL 300
Db 241 LAGMRPIMLELGGKDGAGIVLADADLNAKQIVAGAYDSGORCTAIKRVLVVEEVADEL 300

QY 301 AEKISENVAKLSVGDPPDNATVTPVIDNSADFTESLVDARQKAGELNEFKRDGRLLT 360
Db 301 AEKISENVAKLSVGDPPDNATVTPVIDNSADFTESLVDARQKAGELNEFKRDGRLLT 360

QY 361 PGLFDRVTLDMKLAWEEPFGPILPIIRVKDAEAVAIAANKSDFGQSSVFTDFOKAFDI 420
Db 361 PGLFDRVTLDMKLAWEEPFGPILPIIRVKDAEAVAIAANKSDFGQSSVFTDFOKAFDI 420

QY 421 ANKLEVGTVHNNKTGRGPDNFPFLGLKSGAGVQGIIRYSIEAMTNVKSIVLDMK 475
Db 421 ANKLEVGTVHNNKTGRGPDNFPFLGLKSGAGVQGIIRYSIEAMTNVKSIVLDMK 475

RESULT 2

Q8E0B1 PRELIMINARY; PRT; 475 AA.

ID Q8E0B1 AC Q8E0B1; DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent.
GN GAPN OR SAG0823.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547;
RA Tettelin H., Masiognani V., Cielesiewicz M.J., Eisen J.A., Peterson S.,
RA Wessels L.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kollan J.F., Madupu R., Lewis M.R.,
RA Radu D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014229; AM99710.1; -;
DR TIGR; SAG0823; -;
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.

KW Complete proteome.
SQ SEQUENCE 475 AA; 51218 MW; F91C086696BAE668 CRC64;

Query Match 99.5%; Score 2382; DB 16; Length 475;
Best Local Similarity 99.2%; Pred. No. 2.8e-139;
Matches 471; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKKEYQNVNGEWSKSSVNOIEILSPIDSSISLGFVPAMTREVVDHAKAGREALPAAALT 60
Db 1 MTKKEYQNVNGEWSKSSVNOIEILSPIDSSISLGFVPAMTREVVDHAKAGREALPAAALT 60

QY 61 VYERAQYLHKAADIIERDKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120
Db 61 VYERAQYLHKAADIIERDKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120

QY 121 SADEGKMDASTGHKLAVTRPPQVIGVLAIPNYPNLSGSKIAPALIGNVVMFKPPT 180
Db 121 SADEGKMDASTGHKLAVTRPPQVIGVLAIPNYPNLSGSKIAPALIGNVVMFKPPT 180

QY 181 QGSVSLVLAKAFAPAGLPGVFNITGRGSEIGDYIVHEEVNFNFTGSTPVQRIK 240
Db 181 QGSVSLVLAKAFAPAGLPGVFNITGRGSEIGDYIVHEEVNFNFTGSTPVQRIK 240

QY 241 LAGMRPIMLELGGKDGAGIVLADADLNAKQIVAGAYDSGORCTAIKRVLVVEEVADEL 300
Db 241 LAGMRPIMLELGGKDGAGIVLADADLNAKQIVAGAYDSGORCTAIKRVLVVEEVADEL 300

QY 301 AEKISENVAKLSVGDPPDNATVTPVIDNSADFTESLVDARQKAGELNEFKRDGRLLT 360
Db 301 AEKISENVAKLSVGDPPDNATVTPVIDNSADFTESLVDARQKAGELNEFKRDGRLLT 360

QY 361 PGLFDRVTLDMKLAWEEPFGPILPIIRVKDAEAVAIAANKSDFGQSSVFTDFOKAFDI 420
Db 361 PGLFDRVTLDMKLAWEEPFGPILPIIRVKDAEAVAIAANKSDFGQSSVFTDFOKAFDI 420

QY 421 ANKLEVGTVHNNKTGRGPDNFPFLGLKSGAGVQGIIRYSIEAMTNVKSIVLDMK 475
Db 421 ANKLEVGTVHNNKTGRGPDNFPFLGLKSGAGVQGIIRYSIEAMTNVKSIVLDMK 475

RESULT 3

Q8E5Y8 PRELIMINARY; PRT; 475 AA.

ID Q8E5Y8 AC Q8E5Y8; DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN GBS0841.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM916 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaeser P., Rusniok C., Buchrieser C., Chevalier F., Prangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766847; CAD46485.1; -;
DR SagalList; gbs0841; -;
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 475 AA; 51217 MW; DE33AE79B81A66B6 CRC64;

Query Match 99.4%; Score 2379; DB 16; Length 475;
 Best Local Similarity 99.2%; Pred. No. 4.3e-139;
 Matches 471; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LKEYQYNYNGEWSKSNVQIEILSPIDSSLGFPVPMATREEDVHAKGREALPAWAAALT 60
 DB 1 MKEYQYNYNGEWSKSNVQIEILSPIDSSLGFPVPMATREEDVHAKGREALPAWAAALT 60

QY 61 VYERAOYLKAAADIIERDKKEIATVLAKEISKAYNASVTEVVTADLIRYAAEGIRLST 120
 DB 61 VYERAOYLKAAADIIERDKKEIATVLAKEISKAYNASVTEVVTADLIRYAAEGIRLST 120

QY 121 SADEGGKMDASTGHKLAVIRROPGVIGVLAIPYNNVNLSSGKIAPALIGNVVMPKPT 180
 DB 121 SADEGGKMDASTGHKLAVIRROPGVIGVLAIPYNNVNLSSGKIAPALIGNVVMPKPT 180

QY 181 QGSVGLVLAFAFAEAGLPAGVNTITGRGSEIGDYIIEHEEVNFINTGTPVGRIGK 240
 DB 181 QGSVGLVLAFAFAEAGLPAGVNTITGRGSEIGDYIIEHEEVNFINTGTPVGRIGK 240

QY 241 LAGMRPTMLELGGKDGAGVLADADLONAAKQIVAGAYDYGORCTAIKRVLVVEEVADEL 300
 DB 241 LAGMRPTMLELGGKDGAGVLADADLONAAKQIVAGAYDYGORCTAIKRVLVVEEVADEL 300

QY 301 AEKISENVAKLSVGDPPDNATVTPVIDDNSADFTESLVVDARQKAKELNEFKRDGRLLT 360
 DB 301 AEKISENVAKLSVGDPPDNATVTPVIDDNSADFTESLVVDARQKAKELNEFKRDGRLLT 360

QY 361 PGLFDHVTLDMLKAWBEPFGPILPIIRVKDAEEVAJANKSPGLQSSVTRDFQKAFDI 420
 DB 361 PGLFDHVTLDMLKAWBEPFGPILPIIRVKDAEEVAJANKSPGLQSSVTRDFQKAFDI 420

QY 421 ANKLEVTGHVHNKTKRGPNFPFLGLKSGAGVGQIRYSIEAMTNVKSIVLDMK 475
 DB 421 ANKLEVTGHVHNKTKRGPNFPFLGLKSGAGVGQIRYSIEAMTNVKSIVLDMK 475

RESULT 4
 Q97QT0 PRELIMINARY; PRT: 474 AA.

ID Q97QT0 (TrEMBLrel. 18, Created)
 AC Q97QT0 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent.
 GN SP1119.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Haft D.H., Dodson R.J.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Nelson W.C., Peterson J.D.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Radune D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Hansen C.L.,
 RA Holzhapple E., Khouri H., Wolf A.M., Utterback T.R., Hickey E.K.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae".
 RL Science 293:498-506 (2001).
 DR EMBL; AB007413; AAK5230.1; -.
 DR PIR; E95129; E95129.
 DR TIGR; SP1119; -.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldenhyde dehydr.
 DR Pfam; PF00171; aldedh; 1.

DR PROSITE; PS00070; ALDEHYDE DEHYDR. CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR. GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 474 AA; 51127 MW; 89D8C05F02AC0048 CRC64;

Query Match 75.7%; Score 1812; DB 16; Length 474;
 Best Local Similarity 74.1%; Pred. No. 5.4e-104;
 Matches 349; Conservative 57; Mismatches 65; Indels 0; Gaps 0;

QY 5 YQYVNGEWSKSNVQIEILSPIDSSLGFPVPMATREEDVHAKGREALPAWAAALT 64
 DB 4 YQNLVNGKSSSQEITTSYIPNQELGTVPMTQTEADEAMQAAALPAWRAALSAVER 63

QY 65 AQYLKKAADIIERDKKEIATVLAKEISKAYNASVTEVVTADLIRYAAEGIRLST 124
 DB 64 AAYLHKTAALIERDKKEIATVLAKEISKAYNASVTEVVTADLIRYAAEGIRLST 123

QY 125 GKGMDASTGHKLAVIRROPGVIGVLAIPYNNVNLSSGKIAPALIGNVVMPKPTQGSV 184
 DB 124 GGSFEATSKKLAVIRROPGVIGVLAIPYNNVNLSSGKIAPALIGNVVMPKPTQSSI 183

QY 185 SGLVLAFAFAEAGLPAGVNTITGRGSEIGDYIIEHEEVNFINTGTPVGRIGK 244
 DB 184 SGLVLAFAFAEAGLPAGVNTITGRGSEIGDYIIEHEEVNFINTGTPVGRIGK 243

QY 245 REIMLELGGKDGAGVLADADLONAAKQIVAGAYDYGORCTAIKRVLVVEEVADEL 304
 DB 244 REIMLELGGKDGAGVLADADLONAAKQIVAGAYDYGORCTAIKRVLVVEEVADEL 303

QY 305 SENVAKLSVGDPPDNATVTPVIDDNSADFTESLVVDARQKAKELNEFKRDGRLLT 364
 DB 304 QEVSKLTGVDPPDNATVTPVIDDNSADFTESLVVDARQKAKELNEFKRDGRLLT 363

QY 365 DRVTLDMKAWBEPFGPILPIIRVKDAEEVAJANKSPGLQSSVTRDFQKAFDI 424
 DB 364 DQVTKMDKAWBEPFGPILPIIRVKDAEEVAJANKSPGLQSSVTRDFQKAFDI 423

QY 425 EVGTGHVHNKTKRGPNFPFLGLKSGAGVGQIRYSIEAMTNVKSIVLDMK 475
 DB 424 EVGTGHVHNKTKRGPNFPFLGLKSGAGVGQIRYSIEAMTNVKSIVLDMK 474

RESULT 5
 Q8DPS7 PRELIMINARY; PRT: 474 AA.

ID Q8DPS7 (TrEMBLrel. 23, Created)
 AC Q8DPS7 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.9)
 GN GAPN OR SPR1028.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E., P.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
 RA Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6".
 RL J. Bacteriol. 193:5709-5717 (2001).
 DR EMBL; AE008475; AAK99832.1; -.
 DR PIR; D98000; D98000.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.

DR GO: 0008152; P:metabolism; IEA.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE DEHYDR CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR GLU; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 474 AA; 51076 MW; 87BC0A8BBD7E5E21 CRC64;

Query Match 75.3%; Score 1802; DB 16; Length 474;
 Best local similarity 73.7%; Pred. No. 2.2e-103;
 Matches 347; Conservative 58; Mismatches 66; Indels 0; Gaps 0;

QY 5 YQNYNGEKKSSVNOIETLSPIDSSIGFVPAMTRFEEVDHAKGAGREALPAWALTVER 64
 DB 4 YQNLVNGKSSQEEITYSPIQNOBELGTVPAMTQTEADEAMQARAALPAWRALSAIER 63
 QY 65 AYLHKAADIIRDEKEETATVLAKESIKAYNASVTEVVRTADLIRVAABEGIRLSTGAE 124
 DB 64 AYLHKAADIIRDEKEETATVLAKESIKAYNASVTEVVRTADLIRVAABEGIRLSTGAE 123
 QY 125 GQMDASTGHKLAVIRROPVGIIVLAIPYNPVNLGSKIAPALIGNVVMKPPPTQGSV 184
 DB 124 GGGFPAASKNKLAVRREPVGIVLAIPYNPVNLGSKIAPALIGNVVMKPPPTQGSV 183
 QY 185 SGLVLAKAFEAAGIPAGVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGVORIGKLAGM 244
 DB 184 SGLVLAKAFEAAGIPAGVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGVORIGKLAGM 243
 QY 245 RPIMLELGGKAGIIVLADADLNAAKQIVAGAYDSGORCTAIKRVLVVEEVADELAEKI 304
 DB 244 RPIMLELGGKAGIIVLADADLNAAKQIVAGAYDSGORCTAIKRVLVVEEVADELAEKI 303
 QY 305 SENVAKUSVGDPPDNATVTPVIDNSADFTESLVVDARQKAKELNEFKRDLITPGLF 364
 DB 304 QEEVSKLTVDGPPDNADITPPVIDNSADFTESLVVDARQKAKELNEFKRDLITPGLF 363
 QY 365 DHVTLDMKLAWEEPFGPIPIIRVKDAEVAALANKSDFGIQSSVTRDFQKAFDI 424
 DB 364 DQVTKDMKAWEEPFGPIPIIRVKDAEVAALANKSDFGIQSSVTRDFQKAFDI 423
 QY 425 EYGVTHNNKTGRGPNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 475
 DB 424 EYGVTHNNKTGRGPNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 474

RESULT 6
 Q8P0F5
 ID Q8P0F5 PRELIMINARY; PRT; 475 AA.
 AC Q8P0F5;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Putative NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
 GN GAPN OR SPY18.1383.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Scudervant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 group A Streptococcus strains associated with acute rheumatic fever
 outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
 DR EMBL; AE010058; AAL9798.1; -;
 DR GO: 0016491; F:oxidoreductase activity; IEA.
 DR GO: 0008152; P:metabolism; IEA.

DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE DEHYDR CYS; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 475 AA; 50381 MW; 335E29B77212DFC5 CRC64;

Query Match 68.7%; Score 1645; DB 16; Length 475;
 Best local similarity 65.9%; Pred. No. 1.2e-93;
 Matches 313; Conservative 75; Mismatches 87; Indels 0; Gaps 0;

QY 1 LTKQYNYNGEKKSSVNOIETLSPIDSSIGFVPAMTRFEEVDHAKGAGREALPAWALT 60
 DB 1 MAKQYXNLVNGEKKSSVNOIETLSPIDSSIGFVPAMTRFEEVDHAKGAGREALPAWALT 60
 QY 61 YVERAQYLHKAADIIRDEKEETATVLAKESIKAYNASVTEVVRTADLIRVAABEGIRLST 120
 DB 61 YVERAQYLHKAADIIRDEKEETATVLAKESIKAYNASVTEVVRTADLIRVAABEGIRLST 120
 QY 121 SADEGGKNDASTGHKLAVIRROPVGIIVLAIPYNPVNLGSKIAPALIGNVVMKPPPT 180
 DB 121 EVLEGGSTFEAKSKKATVRRPEVGLVLAIPYNPVNLGSKIAPALIGNVVMKPPPT 180
 QY 181 QGSVSGLVLAFAFAEAGIPAGVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGVORIGK 240
 DB 181 QGSVSGLVLAFAFAEAGIPAGVNTITGRGSEIGDYIVHEEVEVNFINTGSTPGVORIGK 240
 QY 241 LAGMRPTMLELGGKAGIIVLADADLNAAKQIVAGAYDSGORCTAIKRVLVVEEVADEL 300
 DB 241 LAGMRPTMLELGGKAGIIVLADADLNAAKQIVAGAYDSGORCTAIKRVLVVEEVADEL 300
 QY 301 AEKISENVAKUSVGDPPDNATVTPVIDNSADFTESLVVDARQKAKELNEFKRDLIT 360
 DB 301 AEKISENVAKUSVGDPPDNATVTPVIDNSADFTESLVVDARQKAKELNEFKRDLIT 360
 QY 361 PGIFDHVTLDMKLAWEEPFGPIPIIRVKDAEVAALANKSDFGIQSSVTRDFQKAFDI 420
 DB 361 PGIFDHVTLDMKLAWEEPFGPIPIIRVKDAEVAALANKSDFGIQSSVTRDFQKAFDI 420
 QY 421 ANKLEVGTVHNNKTGRGPNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 475
 DB 421 ANKLEVGTVHNNKTGRGPNPFPLGLKSGAGVQGIKYSIEAMTNVKSIVLDMK 475

RESULT 7
 Q99Z67
 ID Q99Z67 PRELIMINARY; PRT; 475 AA.
 AC Q99Z67;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Putative NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
 DE (EC 1.2.1.9).
 GN GAPN OR SPY1371.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
 DR EMBL; AE006575; AAK34198.1; -;
 DR HSRF; P51977; 1BX5.
 DR GO: 0008886; F:glyceraldehyde-3-phosphate dehydrogenase (N. . .); IEA.
 DR GO: 0016491; F:oxidoreductase activity; IEA.
 DR GO: 0008152; P:metabolism; IEA.
 DR InterPro: IPR002086; Aldehyde_dehydr.

[illegible]

RESULT 9	
Q97D25	PRELIMINARY; PRT; 482 AA.
AC	Q97D25
DT	01-OCT-2001 (TREMBlrel. 18, Created)
DT	01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT	01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
GN	CAC3657.
OS	Clostridium acetobutylicum.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
NC	NCBI TaxID=1488;
EN	[1]
RP	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 824 / DSM 792 / VOM B-1787;
RX	MEDLINE=2135925; PubMed=1466286;
RA	Goellling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA	Nielsen R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA	Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA	Bennett G.N., Koonin E.V., Smith D.R.;
RT	"Genome sequence and comparative analysis of the solvent-producing
RT	bacterium Clostridium acetobutylicum.";
RT	J. Bacteriol. 183:4823-4838 (2001).
DR	EMBL; AE007861; AAK81579.1; -.
DR	PIR; H97348; H97348.
DR	GO; GO:0016491; P:oxido-reductase activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPR002086; Aldehyde_dehydr.
DR	Pfam; PF00171; aldedh; 1.
DR	PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR	PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW	Complete proteome.

RESULT 8	
Q8K707	PRELIMINARY; PRT; 467 AA.
ID	
Q8K707	
AC	
DT	01-OCT-2002 (TREMBLrel. 22, Created)
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE	Putative NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
GN	GAPN OR SPYN3_1045.
OS	Streptococcus pyogenes (serotype M3).
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
NCBI_TaxID=198466;	
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=MGAS315 / Serotype M3;
RC	MEDLINE=22133808; PubMed=12122206;
RX	Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,
RA	Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA	Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA	Schlievert P.M., Wuessler J.M.;
RT	"Genome sequence of a serotype M3 strain of group A Streptococcus:
RT	phage-encoded toxins, the high-virulence phenotype, and clone
RT	emergence.";
RT	Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RL	EMBL; AE014157; AA079652.1; --
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPR002086; Aldehyde_dehyd.
DR	Pfam; PF001171; aldehyd. 1.
DR	PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS: 1.

[illegible]

RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,
 RA Tracy S.E., Banh J., Bowser L., Carninci P., Chung M.K.,
 RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L.,
 RA Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
 RA Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,
 RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
 RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
 RA Kawai J., Kim C.J., Narusaka M., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Tang C.C., Toriumi M., Wallender E.K., Wong C.,
 RA Wu H.C., Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J.,
 RA Theologis A., Davis R.W.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RL "Arabidopsis ORF clones";
 RT Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 RN [6]
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR EMBL; AC005967; AAD03388.1; -;
 DR EMBL; AY037205; AAK59790.1; -;
 DR EMBL; AY136409; AAO97075.1; -;
 DR EMBL; BT004551; AAO42797.1; -;
 DR PIR; F84634; F84634.
 DR HSSP; F51977; 1BX5.
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 496 AA; 53060 MW; 9231656A951175D5 CRC64;

Query Match 48.8%; Score 1168.5; DB 10; Length 496;
 Best Local Similarity 50.3%; Pred. No. 3.9e-64;
 Matches 240; Conservative 84; Mismatches 142; Indels 11; Gaps 5;
 QY 5 YQYVNGEWK--SSVNGIETLSPIDSSLGFPVPAWTRREVDHAKAGREALPAWALTYY 62
 DB 16 KYKADGKWTSSSGSKVAIMPATPKTQYKQVQCTQEEVNVAVMELAKSKQSWAKTPIW 75
 QY 63 ERAQYLKAAADIIEERKEEATVLAKESKAYNASVTEVVRTADLIRYAAEEGIRLSTSA 122
 DB 76 KRAELHKAALAKONKPAESLVKEIAKPAKDSVTEVVRSGDLISYCAEEGVRI--L 132
 QY 123 DEGKM--DASTGH--KLAVTRQPGVILVATPNYPNLSGSKIAPALIGNVWFK 177
 DB 133 GEGKFLSDSPFGNDRTKYCLTSKIPGLVLAIPPFPNPNLAVSKIPALIGNSLVLK 192
 QY 178 PPTQGSVGLVAKAFEAAGLPAGVNTITGRGSEIGDYIVEHEEVNFNFTGTSVQGR 237
 DB 193 PPTQGAVALHWHCHFLAGFPKGLISCITGKSGEIGDPLTHPVAVNCISFTGG-DTGI 251

QY 238 IGKLAGNREIMLELGGKDGIVLADADLNAAKOIVAGAYDYSQGRCTAIRKVLWEEVA 297
 DB 252 ISKKAGMIPLQWELGGKDACIVLDDADLDLVASNIKGGFYSQGRCTAVKVLWESVA 311
 QY 298 DELAEKISNVAKLSVGDPFDNATVTPVIDNSADPFIESLVVDARQKAGKELNEFKDGR 357
 DB 312 DELVEKVKAKVAKLTGVPPENSDITAVVSESSANFTIEGLWMDAKEKGATFCQYKREGN 371
 QY 358 LLPGLFDHVTLMKLAWEPPGPIPIIRVKDAEEAVALANKSPDFGLQSSVFTTRDFOKA 417
 DB 372 LIWELLNDRPDMRIAWEBPGFVVPVIRINSVEGINHCNASNFGLOGCVFYDKINKA 431
 QY 418 FDIANKLEVGTVHNNKTRGPDNFPFLGLKSGAGVQGIKIRYSIEAMTNVKSIVLDM 474
 DB 432 ILLSDAMETGTQVINSAPARGPDHPFPFGLKDSGSGQVNTSINLMTKVKTVINL 488
 RESULT 16
 Q8S4Y9 PRELIMINARY; PRT; 507 AA.
 ID Q8S4Y9
 AC Q8S4Y9;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Glyceraldehyde-3-phosphate dehydrogenase.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pillai A., Akiyama T.;
 RT "Flood induced expression of glyceraldehyde-3-P dehydrogenase in rice";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR EMBL; AF357884; AAM00227.1; -;
 DR Gramine; Q8S4Y9; -;
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 507 AA; 54207 MW; 15B7C6090463FFD5 CRC64;

Query Match 47.9%; Score 1145.5; DB 10; Length 507;
 Best Local Similarity 48.4%; Pred. No. 1.1e-62;
 Matches 235; Conservative 88; Mismatches 142; Indels 21; Gaps 7;
 QY 5 YQYVNGEWKSSVN--QIEILSPIDSSLGFPVPAWTRREVDHAKAGREALPAWALTYY 62
 DB 19 YRYVADGEWRVRSAGSKSVAIVNPTTTLTYRQVQCTQEEVNVKMTAKVAKWARTPLW 78
 QY 63 ERAQYLKAAADIIEERKEEATVLAKESKAYNASVTEVVRTADLIRYAAEEGIRLSTSA 122
 DB 79 KRAELHKAALAKELHKHTPIAECLNKEIAKPAKDAISEVVRSGDLVSYTAEEGVRIILGE- 137
 QY 123 DEGKM--DASTG---HKAVTRQPGVILVATPNYPNLSGSKIAPALIGNVWFK 176
 DB 138 ---GKLVSDSPGNERNKYCLSSKPLGVLAIPPFPNPNLAVSKIPALIGNALVL 194
 QY 177 KPPTQGSVGLVAKAFEAAGLPAGVNTITGRGSEIGDYIVEHEEVNFNFTGTSVQGR 236
 DB 195 KPPTQGAVALHWHCHFLAGFPKGLINCITGKSGEIGDPLTHPVAVNCISFTGG-DTGI 253
 QY 237 RIGKLAGMPEIMLELGGKDGIVLADADLNAAKOIVAGAYDYSQGRCTAIRKVLWEEVA 296
 DB 254 AISKKAGMIPLQWELGGKDACIVLDDADLDLVASNIKGGFYSQGRCTAVKVLWESV 313
 QY 237 ADELAEKISNVAKLSVGDPFDNATVTPVIDNSADPFIESLVVDARQKGA---KE----- 348

[illegible]

RESULT 17	PRELIMINARY;	PRT;	508 AA.
Q84PC4			
ID			
AC	Q84PC4		
DT	01-JUN-2003 (TRENBUrel. 24, Created)		
DT	01-JUN-2003 (TRENBUrel. 24, Last sequence update)		
DT	01-OCT-2003 (TRENBUrel. 25, Last annotation update)		
DE	NADH-dependent glyceraldehyde-3-phosphate dehydrogenase (Fragment).		
OS	Oryza sativa (japonica cultivar-group).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzeae; Oryza.		
NCBI_TaxID=39947;			
[1]			
RP	SEQUENCE FROM N.A.		
RN	STRAIN=22820;		
RC	MEDLINE=22584437; PubMed=12684538;		
RA	Cooper B., Clarke J., Buchworth P., Kress J., Hutchison D., Park S.,		
RA	Gumil S., Dunn M., Luginbuhl P., Ellero C., Goff S.A., Glazebrook J.;		
RT	"A network of rice genes associated with stress response and seed		
RT	development.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 100:4945-4950(2003).		
DR	EMBL: AY224439; AA072558.1; "		
DR	GO: GO:0016491; F-oxidoreductase activity; IEA.		
DR	GO: GO:0008152; P-metabolism; IEA.		
DR	IncerFpro; IPR020086; Aldheyde_dehydr.		
DR	Ffam; PF00171; aldedch; 1.		
DR	PROSITE; P800070; ALDSHYDE_DEHYDR_CYS; 1.		
DR	PROSITE; P800687; ALDSHYDE_DEHYDR_GLU; 1.		
NON TER	1		
SEQUENCE	508 AA; 54082 MW; FF02YACBF2743F58 CRC64;		

QY	303	KISENVAKLSVGDPDPNATVTPVIDNDSADFTESIVVDARQKAKELNEFRDGRLLITPG	362
Db	329	KYKAKLAKLVGPPADSDIPVTVTESSANFTIEGLVMDAKEGATFCQEYRREGNLTWPL	388
QY	363	LFSDHVTLDMLKMEWEPGPIIPIIRVKDAAEVAJANKSDFGLSQSVFTRDFQKAPDIAN	422
Db	389	LIDHVRPDMRIAMWEPFGVLPVIRINSVEBGHHHCNASFGLQCGCVTFKDKINXALMISD	448
QY	423	KLEVGTVHNNKTKGRGPONFFFLGLKGSGVQGGIRXYSIEAMTNVKSIVLDM	474
Db	449	ANETCTQVINSARGPDHFFPQGLKXDSGLGSGQGITNSINMWTKYKSVINL	500

RESULT 18	
Q8EV79	
ID	Q8EV79 PRELIMINARY; PRT; 478 AA.
AC	Q8EV79;
DC	Q8EV79;
DT	C1-MAR-2003 (TrenBurel. 23, Created)
DT	C1-MAR-2003 (TrenBurel. 23, Last sequence update)
DT	C1-OCT-2003 (TrenBurel. 25, Last annotation update)
DE	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
GN	MYPE4710.
OS	Mycoplasma penetrans.
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=28227;	
[1]	SEQUENCE FROM N.A.
RN	STRAIN=HP-2.
RC	MEDLINE=22354719; PubMed=12465555;
RX	Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K., Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RA	"The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans.";
RL	Nucleic Acids Res. 30:5293-5300(2002).
DR	EMBL; AP004172; BAC4260.1; -
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPR002086; Aldehyde dehydr.
DR	InterPro; IPR002332; PII_GlnB_UMP_s.
DR	Pfam; PF00171; aldedh_1-
DR	PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
DR	PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
DR	PROSITE; PS00496; PII_GlnB_UMP; 1.
KW	Complete proteome.
SEQ	SEQUENCE 478 AA; 52311 MW; AD0A6110273087C5 CRC64;

Query Match	43.8%;	Score 1047.5;	DB 16;	Length 478;
Best Local Similarity	46.5%;	Pred.No 1.1e-56;		
Matches 218;	Conservative 90;	Mismatched 152;	Indels 9;	Gaps 3

QY	8	YUNGFWKSVUNQTEILSLPTDDSSIGFVPAMTPEEVVDHAWKAGREALPAWAALTVYERAOY	67
Db	11	YINGFVDSEKTIPIVSTIDGSTTGSVAALSKKDLDMAFEGAHLAFNUMRNLTPEMEIKK	70
QY	68	LHKAADIIRDKKEIATVLAKEISKAYNASVTEVVRT-----ADLIRYAAEGIRLSAD	123
Db	71	IKFEAFYIAEKEFLATLMSVEIGKSYNDALKEVRSYENIFETIEWEKEFVNPVSIGP	130
QY	124	EGKMDASTGHKLAVIRQPVGIVLAIAPNVPVNLSSKIAPALIGNVWVFPPTQGS	183
Db	131	EVNKKIGKTGYFVNV-----PVGWVLAIPFNPINLGIAKIITPLLVAGNTVVFPAQTGS	186
QY	184	VSGLVLAKAFBAEAGLPAGVFNITTCRGESEIGYIVVEHEVNFNFTGSTPVGORIGKL	243
Db	187	LVSSQLAKYFDQANFTAGVFNLTGKGESEIGYIILENKRQQATFTFGSDTIGFKLAKLP	246
QY	244	MRPIMLEJGGKDAGIVLADALDNAAKQIVAGADYSGQRCTAIKRVLVEEVADELAEX	303
Db	247	MKPLVLEJGGKORAIYVNADVELAAKEIIKGAFYSGQRCTAIKRVLTVEIAADLLI	306
QY	304	IGENVAKLSVGDFDNATVTPVIDNSADFTESLVVDAROKGAKELNEFEKDG-RLITPG	362
Db	307	LIXEAKKLKVGPNLNNPTDPLIKRAVDNFNTDLDDAINHGVALVYGGNGYNNLRHT	366

QY	363	LFDHVTLDMKLANEBEPFGPLPIIRVDAAEAVATANKSDFGLQSFTRFQKAPDIAN	422
		: : : :	:
Db	367	ILDNVSTKSALKEAEEFPFGLPIIRVTINEIAIANASEYGLQGSVFTKOLEARTIAK	426
		: : : :	:
QY	423	KLEVGTTHNNKTGRGPDNPFPLGLKGSGAGVGQIRYSIEAMTNVKISIV	471
		: : : :	:
Db	427	YLDTGVNIINRGSRGPDIFPFIIGIKSGFGVGQIRDSLAKMTRVGIV	475
		: : : :	:
 RESULT 19 Q8LK61			
ID	Q8LK61	PRELIMINARY;	PRT; 496 AA.
AC	Q8LK61;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase		
DE	(EC 1.2.1.9).		
OS	Triticum aestivum (Wheat).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta, Liliopsida; Poales; Poaceae; Pooidae;		
OC	Triticeae; Triticum.		
OX	NCBI_TaxID=4565;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Leaf;		
RA	Bustos D.M., Iglesias A.A.;		
RL	Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.		
CC	- ~ SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.		
DR	EMBL; AF5211191; XAM77679.1; -		
DR	GO; GO:0008886; F:glyceraldehyde-3-phosphate dehydrogenase (N. . ; IEA.		
DR	GO; GO:0016491; P:oxidoreductase activity; IEA.		
DR	GO; GO:0008152; P:metabolism; IEA.		
DR	InterPro; IPR002086; Aldehyde_dehydr.		
DR	Ffam; PF00171; aldehyd; i.		
DR	PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.		
DR	Oxidoreductase.		
SW	SEQUENCE 496 AA; 52872 MW; 20F9D07F1F840FFB CRC64;		

Db 431 AIMSADAMESGTVQINSAPAGDPHPFQGLKDSGSGQITNSINMTWKSTVINL 488

RESULT 20

ID	Q8L5J9	PRELIMINARY;	PRT;	496 AA.
AC	Q8L5J9;			
AD	Q8L5J9;			
DT	01-OCT-2002 (TRENBLrel. 22, Created)			
DT	01-OCT-2002 (TRENBLrel. 22, Last sequence update)			
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE	Nonphosphorylating glyceralddehyde-3-phosphate dehydrogenase (EC 1.2.1.9)			
DE	Triticum aestivum (Wheat).			
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;			
OC	Triticeae; Triticum.			
OC	NCBI TaxID:4565;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endosperm;			
RC	Eusot D.M., Iglesias A.A.;			
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.			
RL	-!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.			
CC	EMBL; AF5211190; AA071678.1; -.			
DR	GO; GO:0008886; F:glyceraldehyde-3-phosphate dehydrogenase (N. . ; IEA.			
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.			
DR	GO; GO:0008152; P:metabolism; IEA.			
DR	InterPro; IPR02086; Aldehyde_dehydr.			
DR	Pfam; PF00171; aldedh; 1.			
DR	PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.			
KW	Oxidoreductase.			
KW	SEQUENCE 496 AA; 52869 MW; 8C1B1F15FF33881B CRC64;			
Qy	Query Match 42.9%; Score 1026.5; DB 10; Length 496;			
Db	Best Local Similarity 45.0%; Pred.No. 2.4e-55;			
Db	Matches 215; Conservative 78; Mismatches 172; Indels 13; Gaps 4;			
Qy	5 YQNYNGEWSKSYN--QIEILSPDSSSLGFPVPMTEEYDHAKAGREALPAWALTYY 62			
Db	16 YKYADGWRASAGTKVTAIVNPTTQTVRVQACTGEEVKNVMDAAKVAKKWARTPLM 75			
Qy	63 ERAQLHKAADIIERKEETATVLAKEISKAYNASVTEVVRTDLIRYAPEGRIUSTA 122			
Db	76 KRAELHKAAILKEHKHTPIACLVKEIAKPAKDAVSEVRSGLVSYTAEGRVILGE- 134			
Qy	123 DEGGKMDASTG-----HKLAVIRQPVGLVLAIPNYPNISGSKIAPALIGNVNWE 176			
Db	135 ---GKLVSBSFFGNERNKYCLSSKPLGVLAIPFPNYPNANLAGSKIPALTAGNALVL 191			
Qy	177 KPPTQSGVGLVLAFAFAGLAPAGVNTITQRGSETGDYIVHEEVENFINFTGSTPVQ 236			
Db	192 KPPTQGAVALHWHCFHLGAFPKLLISCVTKGSGELGDLTFHPGVNCTISFTGG-DWGI 250			
Qy	237 RIQKLAGMRPTMLEGGKDAGIIVLADLDNAAKOIVAGAYDSYSGRCRTAIKRVILVEEV 296			
Db	251 AISKAGAVPLXKXKGKDAICVEXADLDLVKANI VKGFGSYSGQRCAKVKVLIIMEXX 310			
Qy	297 ADLAEKISENVAKLSVGGPDFNATVTPVIDDONSADTIESLVDARQKAKELNEFKRDG 356			
Db	311 ADXWZKYNAKLAKVGPPEXXXDITXVVTSSANFTXGSGBCXKGRXNFUPXYQEGK 370			
Qy	357 RLITPGLDFHVTLDMLKAWEPFGPIPIRVKDAEFAVAIANKSDPGLQSSVFTPDFOK 416			
Db	371 XLTWPLLLDHRVDPDXIANXWFFGVLVPIRINSVEGIIHCNASNFGLCGVFTDRINK 430			
Qy	417 AFDIANKEVGTVHINNTKGRPNFFPLGLKSGAGVQGIGRVSYTEAMTNVKSIVLDM 474			
Db	431 AIMSADAMESGTVQINSAPAGDPHPFQGLKDSGSGQITNSINMTWKSTVINL 488			

RESULT 21

99PQC9

DE Aldehyde dehydrogenase.
GN PAE2480.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RA "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AB009879; AAL64225.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 478 AA; 52861 MW; 9F6C9409116223E7 CRC64;
Query Match 32.8%; Score 786; DB 17; Length 478;
Best Local Similarity 37.1%; Pred. No. 1.7e-40;
Matches 176; Conservative 96; Mismatches 192; Indels 10; Gaps 4;
QY 6 QNYVNGWKSNNQIETILSPIDSSGLFVPAMTRBEVDHAKMGREALPAWALTVERA 65
DB 5 KNYINGFIEVEBERPKSPIDGSAIETPVSKREDAKAIDAAYDALKAWQLPAIKRA 64
QY 66 QYLKKAADIIRDEKEIATVLAKEISKAYNASVTBVTADLIRYAEAGIRLSTSADEG 125
DB 65 EYLYKLYEIIKSRSEELINILWEGGGIYKKGWGVFTLRLVNAE-----LARRHYG 118
QY 126 GK-WDASTGHLAVIRPQVIGLAIAPYNYPNVSGSKIAPALIGNVVMKPPTOGSV 184
DB 119 GKTQSDAGVSWMPKPKGVGVITPMNYPUSIMKKIAHALATGNTIVYKASETPV 178
QY 185 SGLVLAKAFAEAGLPAGVFNITGRGSEIGDYIVHEEVNFINTGTPVGQRIKLAG- 243
DB 179 TGFKAEMIHQAGFPKGVNVLVGVSTGVGDEIVNKKVSHVITGESATGREIAKAGG 238
QY 244 -MRPIMLELGGKAGIVLADADLNNAAKQIVAGAYDYGQRCATKRVLVVEVDELA 302
DB 239 ALKTVTLELGGDLIILDADLNLAVRIAVFGAFHQGCTSSKRIIVHVRVYDFVK 298
QY 303 KISENVAKLSVGPDPDNATVT- PVIDDNSADFTESLVVDARQKAKELNEFKRDELITP 361
DB 299 KFEVRSQKLVGDPDRDTVEQGPLISPRQADENEFYQDANVSRGRALTGGKAGAYFWP 358
QY 362 GLFDHVTLDKLAWEPPFGPIIPIRVKDAEAAVAIAANKSDGLQSSVFTTRDFQKAFDIA 421
DB 359 TVFVVDNFRIMREVEFGPIRPVVPVKNDEAVEANDTEYGLSAVVTNINRAFKIA 418
QY 422 NKELVGTWHNNKTRGPDNPFPLGLKSGAGVGIRYSIEMTNVKSIVLDMK 475
DB 419 EAIESGMVHNDVITMLESHVFGGIKASGFRGEGGWSFHETIDRWLTITIR 472
RESULT 23
Q97XA5 PRELIMINARY; PRT; 470 AA.
ID Q97XA5
AC Q97XA5;
DT 01-OCT-2001 (TremBrel. 18, Created)
DT 01-OCT-2001 (TremBrel. 18, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase, NADP dependent (gapN-2)
GN (EC 1.2.1.9).
DE GAPN-2 OR SSO1842.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.

Q9POC9 PRELIMINARY; PRT; 475 AA.
Q9POC9;
DT 01-OCT-2000 (TremBrel. 15, Created)
DT 01-OCT-2000 (TremBrel. 15, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
GN GAPN OR U0362.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Casella G.H.;
RA "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum";
RL Nature 407:757-762(2000).
DR EMBL; AB002133; AAF30771.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 475 AA; 52590 MW; DDB0D06576603752 CRC64;
Query Match 36.3%; Score 868.5; DB 16; Length 475;
Best Local Similarity 39.7%; Pred. No. 1.3e-45;
Matches 190; Conservative 95; Mismatches 177; Indels 17; Gaps 5;
QY 5 YONVNGWKSNNQIETILSPIDSSGLFVPAMTRBEVDHAKMGREALPAWALTVE 63
DB 3 YKTLINGAFVDAKEKLPYVNSNNQIIAYVPHNHEINTEFNHAYLAKFDYTIKY 62
QY 64 RAQYLKKAADIIRDEKEIATVLAKEISKAYNASVTBVTADLIRYAEAGIRLSTSA- 122
DB 63 RCDLLLLKADLDEHKQLAQITSTETAKGLXDSLIEVSADYLRFTVEYQKLMQKPI 122
QY 123 ---DEGGKWDASTGH---KLAVIRPQVIGLAIAPYNYPNVSGSKIAPALIGNVVMF 176
DB 123 IFDE-----TVHHVKNKVAITYRIPGVGLAICFPNYPINLLISKAPALVSGNSLVY 175
QY 177 KPPTQGSVGLVLAKAFAEAGLPAGVFNITGRGSEIGDYIVHEEVNFINTGTPVGQ 236
DB 176 KPSTQGSGLIRISLVEHGVFPKGVNVLVGVSTGVGDEIVNKKVSHVITGESATGREIAKAGG 235
QY 237 RIGKAGMRPIMLELGGKAGIVLADADLNNAAKQIVAGAYDYGQRCATKRVLVVEV 296
DB 236 HIAETISKISLVEELGGKDPALVLDADLNLAVRIAVFGAFHQGCTSSKRIIVHVRVYDFVK 295
QY 297 ADELAKEISENVAKLSVGPDPDNATVT- PVIDDNSADFTESLVVDARQKAKELNEFKRDELITP 361
DB 296 HDLLVNLINKKDALTVGLPQDNPIITPLGNSLNSLNLVDEDAIKKGAIVHOKIYVNE 355
QY 355 DGRLLTGLFDHVTLDKLAWEPPFGPIIPIRVKDAEAAVAIAANKSDGLQSSVFTTRDF 414
DB 356 KNNLLHPLVDNVTKEMRVANEPPFGPIIPITVNSIQEALDILNQSQVGLQACIFITNY 415
QY 415 QKAFDIANKLEVGTWHNNKTRGPDNPFPLGLKSGAGVGIRYSIEMTNVKSIVLDMK 473
DB 416 ASIEQLAQIESGTINKSSRSGFDILPFFGVKDSGFGVGQIVDAILSMITIKGIIN 474
RESULT 22
Q8ZV31 PRELIMINARY; PRT; 478 AA.
ID Q8ZV31
AC Q8ZV31;
DT 01-MAR-2002 (TremBrel. 20, Created)
DT 01-MAR-2002 (TremBrel. 20, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)


```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX  MEDLINE=22220767; PubMed=12235376;
RA  Takami H., Takaki Y., Uchiyama I.;
RT  "Genome sequence of Oceanobacillus theyensis isolated from the Iheya
RT  Ridge and its unexpected adaptive capabilities to extreme
RT  environments.";
RL  Nucleic Acids Res. 30:3927-3935(2002).
DR  EMBL: AP004602; BAC14796.1; -.
DR  GO: GO:0016491; F:oxidoreductase activity; IEA.
DR  GO: GO:0008152; P:metabolism; IEA.
DR  InterPro: IPR002086; Aldehyde dehydr.
DR  Pfam: PF00171; aldedh; 1.
DR  PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR  PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW  Oxidoreductase; Complete proteome.
SQ  SEQUENCE 486 AA; 52280 MW; 39A5A1941EB4013B CRC64;

Query Match      31.4%; Score 750.5; DB 16; Length 486;
Best Local Similarity 35.0%; Pred. No. 2.8e-38;
Matches 171; Conservative 103; Mismatches 186; Indels 29; Gaps 9;

Qy  3 KEYQNYNGWKSQV--NQILSPIDDS--LGFVPAMTRVEVDHAKGREALPANAAL 59
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7 QSYLNFNGRWKSEGGTKEVLPADKNDVVGVIQRTDIDVNEAIAASAKSKEVRL 66
Qy  60 TVYERAOYLHKAADIIRKKEIATVLAKESKAYNASVTEVRTADLIRVAEEGIR-- 117
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 SNERGNYLYKVANILESRLODVAETLTREMGKTLPEAKGETARGVAILRYAGEGLREV 126
Qy  118 --LSTSADEGGKMDASTGKHLAVIRROPVIGVILAIAPYNPVNLGSKIAPALGGNVVM 175
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 GDVIPSTDSSGIWYTI-----RTPLGVGVVITPWNFEVPAIPWKMAPALIVGNTVV 177
Qy  176 FKPTQSGVGLVLAKAFABAGLPAGVNTITGRGSEIGDYIVHEHEVNFNFTGTPVVG 235
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 IKPATEAATAAKWCECFADAEPLDGVVNMVVGSGISVGNRIAEHPDVNGVTFITGSDQTG 237
Qy  236 QRIKLAGMR--PIMLEGGDAGVILADLDNAKQIVAGAYDYSGORCTAIKRLVAV 293
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 KHIGOTALAGAKYQLEMGKRPVIVADDAADLDLAVKATISGLSGTSQKCTATSRVIVQ 297
Qy  294 EVADELAELKISENVAKLSVGDPPDNAT--VTPVIDDNSADFTESLVVDARQKAGKELNEF 352
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
298 RSVYDAFKEKLLTQIQALTVNGLEDGNGWGPSVESQLETVLHYIKKGREGATLLTGG 357
Qy  353 KR-----DGRLLTPGLFDHVTLDMLKWEPPFPILPIIRVKDAEEAVALANKSDPL 405
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
358 HRLTDAHBKDGNEFVTFEDVTDNTMTIAQEEIFGFLVLIKVFETFEAMKLANDTYGL 417
Qy  406 QSSVTRDRDFOKAFDIANKLEVGTWHNNKTRGPDNFPFLGLKSG--AGVQGIRYSIEA 463
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
418 SNAVFTENIKMLTIDIDIEGLVRVNAETAGVQLQAPFGMKSSSSHSREQG--KAAIEF 476
Qy  464 MTNVKSIVL 472
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
477 FTSIKTVFI 485

RESULT 26
Q97XS9 PRELIMINARY; PRT; 470 AA.
AC Q97XS9;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Glyceraldehyde-3-phosphate dehydrogenase, NADP dependent (gapN-1)
DE (NC 1.2.1.9).
GN GAPN-1 OR SS01629.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.

```

```

OX  NCBI_TaxID=2287;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 35092 / DSM 1617 / P2;
RX  MEDLINE=21332296; PubMed=11427726;
RA  She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA  Awayez M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA  De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA  Heikamp-de Jong I., Jaffries A.C., Kozera C.J., Medina N., Peng X.,
RA  Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA  Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA  Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT  "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR  EMBL: AE006776; AAK41843.1; -.
DR  PIR: D90323; D90323.
DR  GO: GO:0008886; F:glyceraldehyde-3-phosphate dehydrogenase (N. .; IEA.
DR  GO: GO:0016491; F:oxidoreductase activity; IEA.
DR  GO: GO:0008152; P:metabolism; IEA.
DR  InterPro: IPR002086; Aldehyde dehydr.
DR  Pfam: PF00171; aldedh; 1.
DR  PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
KW  Oxidoreductase; Complete proteome.
SQ  SEQUENCE 470 AA; 51529 MW; 54BEB97B2D2FFE78 CRC64;

Query Match      31.3%; Score 750; DB 17; Length 470;
Best Local Similarity 36.6%; Pred. No. 2.9e-38;
Matches 171; Conservative 94; Mismatches 186; Indels 16; Gaps 7;

Qy  20 IEILSPIDDSLGFVPAMTRVEVDHAKGREALPANAALTVYERAOYLHKAADIIRDK 79
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7 IEVRSPLNVIGTVRMEKHDEVRGEIEEAYKGFELISKMPLYKRTAILRKVSEILEREQ 66
Qy  80 EETATVLAKETSKAYNASVTEVRTADLIRVAEE--GIRLSTSADEGKMDA-----STG 133
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 ERLAKLLAMEAGKPIRDSRVEVIRASRLFRHAAEEVGVLE---GKNRYVDAYEYPPGNE 123
Qy  134 HKLAVIRROPVIGVILAIAPYNPVNLGSKIAPALGGNVVMFKPTQSGVGLVIAKAF 193
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 NRIVTSVEPIGVVVTAILFPNFPINSFAHKVAPALAGVNVVVKPSINTPLAIEBMKKIL 183
Qy  194 ABAGLPAGVNTITGRGSEIGDYIVHEHEVNFNFTGTPVQGRIGKLA---GMRPIMLE 250
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 VEAGLPDSAVRVVTVGSYSGEIGDEIITHTPLVGLITLTGTTGLKIASKAVSLGKR-IIE 242
Qy  251 LGGDAGVILADLDNAKQIVAGAYDYSGORCTAIKRLVVEEVADELAELKISENVAK 310
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 LGGSDPIILLDANIERASSIAVRARFEYAGQNCNAGKRIIVREIYDKVAFNDKARA 302
Qy  311 LSVGDPFDNAT--VTPVIDDNSADFTESLVVDARQKAKE--LNEFKRDGRLTTPGLFDHV 367
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
303 LKVGDPDLDTTGVGVINKESVENLNSVLADAKAKGKVEILNKGPSGSPFPLTWTNP 362
Qy  368 TLDMLKWEPPFPILPIIRVKDAEEAVALANKSDPLQSSVTFTRDFOKAFDIANKLEV 427
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363 SLDMVLTKSEVFGPIVPIVSVKSDDEAIIANSTANSTGYLQSAIFTNDVNRALKSLRELFG 422
Qy  428 TVHNNKTRGPDNFPFLGLKSGAGVQGRIRYSIEAMTNVKSIVLDM 474
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
423 AVINDSTRULNSLAFGGFKKSGISGREGVRETMLENTENKLIATL 469

RESULT 27
Q8CV96 PRELIMINARY; PRT; 497 AA.
AC Q8CV96;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Aldehyde dehydrogenase.
GN OB0861.
OS Oceanobacillus theyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.

```

```
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus ihneyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RL environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL: AP004596; BAC12817.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 497 AA; 54141 MW; 14F219B301BE54FB CRC64;

Query Match 31.0%; Score 742; DB 16; Length 497;
Best Local Similarity 34.8%; Pred. No. 9,7e-38;
Matches 169; Conservative 98; Mismatches 196; Indels 22; Gaps 6;

QY 2 TKBYQYVNGEWSVNOIEILSPIDSSLGFPVPMTRVEVDHAKGREALPAAALTV 61
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 TDVLKNFIDGWVDVNTTAVTPNANGERIVQVPLSDSHVEEAVOAAKAQKEMALVPA 67

QY 62 YEAQVILKAAADIERKEIATVLAKELSKAYNASVTEVTADLIRYAAEGIRLSTS 121
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 PQBAEVLVGVGMKOKKRLSLTMTENGKVEEAGEVQEGIDMAFYWAGEGRLEF-- 125

QY 122 ADGGGMDASTGKLAIVRQPVGIVLAPVNPVNLGSKIAPALGNNVWFKPPTQ 181
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 ---GQTTPAELKDFMSQRPVGVGIIITPWFPAIATWKSFAIVAGNAVWVPATE 182

QY 182 GSVGLVLAKAFAGLPAAGVFNITGRSGEIGDIYIVHEEVEVNFITGSTPVGRIKL 241
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 TPIMAYELAKIFEAGLPSGVINNVGKSGVGVQHKDIRVIFSTGSDTGRNIASD 242

QY 242 AG--MRPIMLELGGKAGIVLADLDAKQIVAGDYSGQRTAKRVLVVEVADE 299
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 CGQLKKSLENGKGNVIMDDADLDAVEGLWSAFGTSQRTACSRVIVHHVKKT 302

QY 300 LAEKISENVAKLSVGDFF--DNATVPVDDNSADFTESLVVDARQKAKELN----EPK 353
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 LEELLAANDHLITIGNGLEGIKGVPLINEAGLNKIKKIEVKGQEGATLLAGTEWETN 362

QY 354 RGRLLTPGLFHVTLMKLAWEEPFGPIPIIRVKDAEEAVAIANKSDFGLOSSVFTD 413
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 NPGYYFTPTLTDKAPDMRIAQBEIFGPVWSLIEVSFEAEIVNNQVTFGLSSIFTKD 422

QY 414 FQKAFDIANKLEVGVTHNNKTRGPDNFPFLGLKSGS-----AGVGIRYSIEAMTVK 468
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 423 VNRVKAQRLDGLGVIVNAGTTGAETHLPFGTKGTGNHRDMSGV-----ALDVFTEWK 478

QY 469 SIVLD 473
DQ : : : :
Db 479 AIYVD 483

RESULT 28
O27059
ID O27059 PRELIMINARY; PRT; 455 AA.
AC O27059
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
GN MTH978
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
```

```
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000871; AAB85474.1; -.
DR PIR: H69230; H69230.
DR HSSP: P20000; IAG8.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 455 AA; 48825 MW; A98D7BA54AA9C08B CRC64;

Query Match 31.0%; Score 741.5; DB 17; Length 455;
Best Local Similarity 38.2%; Pred. No. 9.1e-38;
Matches 173; Conservative 83; Mismatches 186; Indels 11; Gaps 5;

QY 9 VNGEWSVNOIEILSPIDSSLGFPVPMTRVEVDHAKGREALPAAALTVYERAYL 68
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 IDGEGVSGEKITFVRNPNENGVDVRLPAGRSDEVERAIRAAHEARDAMADLSARKISEKL 64

QY 69 HKAADIIRKEIATVLAKELSKAYNASVTEVTADLIRYAAEGIRLSTSADGGKM 128
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 YDVADELKTELDEFARLITLESKPIRFSRDEVKESVETARLSABEAGRL---YGESIPM 121

QY 129 DASTGKH--LAVIRQPVGIVLAPVNPVNLGSKIAPALGNNVWFKPPTQGSVSG 166
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 DAGTGGKLTGTTRIPGVVAITFPNLALHKGVPALAGNTTILKPSLEAPLSA 191

QY 187 LVLAFAEAGLPAAGVFNITGRSGEIGDIYIVHEEVEVNFITGSTPVGRIKLGRP 246
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 LKLAWILSE--HPAGAVNAVTRGSEVGVVIDSPLDVKITFTGSEVGVGRYISARASMKK 240

QY 247 IMLELGGKAGIVLADLDAKQIVAGDYSGQRTAKRVLVVEVADELAELKISE 306
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 ITLELGGNDPLVMDADIDSAAVEAAVRGSYLVSQVCIARKEIMYHEDVADFADKLVN 300

QY 307 NVAKLSVGDFFDNAT--VTPVDDNSADFTESLVVDARQKAKELNFKRDGRLTTPGLFD 365
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 ITGSLRAGDPMVTRDVGVLNEDAEIEVERVIGRAVEDGALLCGSRRGNFVEPTVLD 360

QY 366 HVTLDMLKAWEEPFGPIPIIRVKDAEEAVAIANKSDFGLOSSVFTDQKAFDIANKLE 425
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 HWFPGMEVIERETFGVSPPIIFENGADAEIRIANGTCVVALQAGVFTENIRIALRMAREIE 420

QY 426 VGTVTHNNKTRGPDNFPFLGLKSGAGVGIR 458
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 AGTVLVNKQSTFRVDHMPF----GDSGAVEVWR 449

RESULT 29
Q8NMB0
ID Q8NMB0 PRELIMINARY; PRT; 484 AA.
AC Q8NMB0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NAD-dependent aldehyde dehydrogenases (EC 1.2.1.-).
GN CGL2668.
```

RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA	Garber R.L., Golltry L., Tolerant E.E., Westbrock-Wadman S., Yuan Y.,
RA	Brody L.I., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA	Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RA	"Complete genome sequence of <i>Pseudomonas aeruginosa</i> PA01, an
RT	opportunistic pathogen.";
RL	Nature 406:959-964(2000). ; -
RL	EMBL; AE004903; AAG08284.1; --
DR	PIR; A83033; A83033.
DR	HSP; P51977; 1BX5.
DR	GO; GO:0016491; P:oxidoreductase activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPRO02086; Aldehyde_dehydr.
DR	Pfam; PF001171; aldehyd; 1.
DR	PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR	PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW	Complete proteome.
SQ	SEQUENCE 489 AA; 52697 MW; DECBA8EGE73793D8 CRC64;

Query Match	31.0%; Score 741.5; DB 16; Length 489;
Best Local Similarity	37.0%; Pred.No.le=37;
Matches 173; Conservative	87; Mismatches 185; Indels 23; Gaps 7

QY	11	GEMK--SWNQIETLPIDGSLGFVPAMTREVDHMKAGREALPAWAALTIVVERAOYL 68
DB	16	GEWRHGCRGLXNPNFPGSLLLEIEQADRDLDAAAYAKAEVQPAWALGCPARAVL 75
QY	69	HKAADIIERKEEIAIVLAKEISKAYNASVEVTATDLIRYAEEGIRLTSA----DE 124
DB	76	YKAVEVFDRHEIRVDWIITRESGS-----TRLKAEIWGAARAITLESASFPARVH 126
QY	125	GGMDASTGHKLAVIBRRQVGIVLIAIPYNYPNLNSGKIAPALIGNVMFKPTQGSV 184
DB	127	GRIVESDPGKERVTRISAIGVGVISFPMNFPHUTQSIAPALALGNVAVVKPASTPV 186
QY	185	-SGLVAKAFEAAGLPAGVFNITTGRGSEIGDYIVVEHEBVNFINTGSTPVGQRIGLAG 243
DB	187	CGILLIARIPEEAAGLEAGLFSSVWGPGSEIGDAFVEHPVPLVTFGTSTFPVRNIGRIAS 246
QY	244	---MRPIMLELGKGAGIVLADLDLNAAKQIVAGADVDSGORCTAIKVLVVEEVADE 299
DB	247	GGAHUKHVALLGNGSNPFVVLGDADLEQAANAAPFGKFLHQGQICMAINRIIVEDSLYDA 306
QY	300	LAEKISENAKLSVGDGP--FDNATVTPVIDDNSADFIESLVWDARQKAKELNEFKRDR 357
DB	307	FRAAFERVRVGLRVGPPQADTA-VGIYVNAQLSGILEKILRARQEGAKPLYEGSVDGQ 365
QY	358	LITPGLFPHVTLDMKLANEEFPGLDIRTVKDAAEVATAANKSDFGLOSSVFTFRDPKA 417
DB	366	LIAPHVFEVTAETMETARDETIFGLVGLLRARDEAAHALEANASEYGS\$SAVFSRDLERA 425
QY	418	FDIANKLRGVTHVNKNKTGRGPDFNPFLGLKSGAGVQGTRYSIEMT 465
DB	426	VRFARQLRAGTHVNDIPWNDEANAPFGEEKSGLGRFNGDMAIBEFT 473

RESULT 31	
Q8U2S5	PRELIMINARY; PRT; 506 AA.
ID Q8U2S5	
AC Q8U2S5;	
DT 01-JUN-2002 (TrEMBLrel. 21, Created)	
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)	
DE Non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase.	
GN PF0755,	
OS Pyrococcus furiosus.	
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;	
OX Pyrococcus	
NCBI_TaxID=2261;	
RN [1]	
RP SEQUENCE FROM N.A.	

RC STRAIN-Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the *Pyrococcus furiosus* genome.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE010193; AL80879.1; -.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR InterPro: IPR02086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 506 AA; 56034 MW; E3B233A26F7A5F42 CRC64;
 Query Match 30.9%; Score 739; DB 17; Length 506;
 Best Local Similarity 34.0%; Pred. No. 1.5e-37;
 Matches 165; Conservative 114; Mismatches 175; Indels 32; Gaps 7;
 QY 4 EYQNYNGEWSVNVQIILSPIDSSSLGFPVPMTRREVDHAM----YAGREALPAAAL 59
 DB 23 EFKTYAGWYGEKFKDKIPKINGEVIARVILSREQIEEALSTTYIGREKIRDPG- 81
 QY 60 TVVERAQYLHKAADIIEERKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLS 119
 DB 82 --EKGISFLKAEEMKAEFDDFVRILLDAGKPKSNMGEVATIE-----RL 129
 QY 120 TSADEGGKM-----DASTGHKLVIRQPGVIGVILAIAPYVNVLSGSKIAPALIG 170
 DB 130 KTFEFGRLIGDYIPGDMSEETLESAIVKREPVGLILAIAPYVNVLSGSKIAPALIG 189
 QY 171 GNVVMPKPTQSSVGLVAKAFABEAGLIPAGVNTITGRSEIGDYIVHEEVNFINFTG 230
 DB 190 GNAVLLKPSADAPLILFTVLELSGLPKESFSLTVPGGFMD-IVRDKIRAITFTG 248
 QY 231 STPGVQIRKLAGMPIMLELGGKAGIVLADADLNAKQIVAGAYDYSQRCATKRV 290
 DB 249 STEVGHIIRTGKAVHLEGGKDAIVINDANIESVEKIVKGIYVSGQRDAIRLI 308
 QY 291 LWEEVDAELAEKISENVAKLSVGDPP--DNATVTPVIDNSADIESLVVDARQKAKE 348
 DB 309 LVEEDIVEQRESVIAQLSKIEPKNPLEDENAINGLIDESANYIEIYKDAIEKGARP 368
 QY 349 LNEFKDGRLLTPGLFD---HVTLDMLKAWBEPGPILPIRVKDAEAVAIANKSDGL 405
 DB 369 LLEFRGRNYVPIVLEVNNEILPTIRAFNEDEFGPIALLIKYKODEAIBANSRPL 428
 QY 406 QSSVFTDFQKAFDIANKLVGTVTHNNKTRGPDNFPFLGLKSGAGVQGIIRYSIEAMT 465
 DB 429 DAAVFTTCKRKARKVAKLVGVSFVINEYPHGIYYPFGKMDSGIGREGIYSLEMLT 488
 QY 466 NVKSIV 471
 DB 489 TTKTIV 494

RESULT 32
 O66573 PRELIMINARY; PRT; 476 AA.
 ID Q59702
 AC Q66573
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Aldehyde dehydrogenase.
 GN ALDH1 OR AQ.186.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VFS;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";
 RL Nature 392:353-358 (1998).
 DR EMBL: AB000677; AAC06525.1; -.
 DR FIR; A70318; A70318.
 DR HSP; P51977; LBXS.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR InterPro: IPR02086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 KW Complete proteome.
 SQ SEQUENCE 476 AA; 52608 MW; CE001282933A9081 CRC64;
 Query Match 30.8%; Score 737; DB 16; Length 476;
 Best Local Similarity 35.0%; Pred. No. 1.9e-37;
 Matches 168; Conservative 96; Mismatches 200; Indels 16; Gaps 6;
 QY 4 EYQNYNGEWSVNVQIILSPIDSSSLGFPVPMTRREVDHAMKAGREALPAAALTYVE 63
 DB 3 EKMLIGGEVWDKEEKIEVITYPYTRKPIGRVPGKDEKVEKAIERAKEGFKEIFSLTAYE 62
 QY 64 RAQYLHKAADIIEERKEEATVLAKEISKAYNASVTEVVRTADLIRYAAEGIRLS---- 119
 DB 63 RYELIMRAAQLKRAEAEFATLVLEVGTIREARTEVQRAIQTLIFSAAEKRVNGETF 122
 QY 120 -TSADEGGKMDASTGHKLVIRQPGVIGVILAIAPYVNVLSGSKIAPALIGNVVMPKP 178
 DB 123 PIDAHNGK-----GKLGFIIRVPVIGVISAITPENFELNSMHKVAPALAAAGNAVILKP 176
 QY 179 PTQSSVGLVAKAFABEAGLIPAGVNTITGRSEIGDYIVHEEVNFINFTGSTPVQORI 238
 DB 177 SERTPLPLMLGSEILLEAGVPPKALSIVFYCG-DVGKAMTHPDVRVVSFTGSRKVGDI 235
 QY 239 GKLAGMPIMLELGGKAGIVLADADLNAKQIVAGAYDYSQRCATKRVLVVEEVAD 298
 DB 236 ARQVIGIKVLELGSNSAILHKGDNLEKAVETVGGIYAGVCISVQRFVHEDLYD 295
 QY 299 ELAEKISENVAKLSVGDPPDNAT-VTPVIDNSADIESLVVDARQKAK-ELAEFK--R 354
 DB 296 EFKELKXRVSKVKGDPDMNEDTDLGPMIAPSELERIQEWISEAVQKAKIEAGLRCAE 355
 QY 355 DORLLTGLFDHVTLDMLKAWBEPGPILPIRVKDAEAVAIANKSDGLSSVFTDF 414
 DB 356 DETFTSFPTVSLVPEDSKLFKEAFAPVAVNPYKIDBEAIRVWNSDYGLQGVFTNDV 415
 QY 415 OKAFDIANKLVGTVTHNNKTRGPDNFPFLGLKSGAGVQGIIRYSIEAMTVKSVILDM 474
 DB 416 KLANCKIEIAGGVLINEGPTFRADMPYGVKYSIGREGPKFAIEDYTEIKTIFDL 475

RESULT 33
 O59702 PRELIMINARY; PRT; 491 AA.
 ID Q59702
 AC Q59702;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE P-hydroxybenzaldehyde dehydrogenase.
 GN PCHA
 OS Pseudomonas putida.
 OG Plasmid pRA4000.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP PARTIAL SEQUENCE FROM N.A.
 RC STRAIN-NCIMB 9866;
 RX MEDLINE=95014081; PubMed=7929007;
 RA Kim J., Fuller J.H., Cecchini G., McIntire W.S.;
 RT "Cloning, sequencing, and expression of the structural genes for the
 RT cytochrome and flavoprotein subunits of p-cresol methylhydroxylase
 RT from two strains of *Pseudomonas putida*.";
 RL J. Bacteriol. 176:6349-6361 (1994).

GN	RA2289.
OS	Bacillus anthracis (strain Ames).
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=198094;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22608414; PubMed=12721629;
RA	Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA	Nelson K.E., Tettelin H., Pouts D.E., Eisen J.A., Gill S.R.,
RA	Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA	Kolman J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA	DeBoy R.T., Madupu R., Dougherty S.C., Durkin A.S., Haft D.H.,
RA	Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA	Berry J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA	Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA	Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.B.,
RA	Fraser C.M.;
RT	"The genome sequence of Bacillus anthracis Ames and comparison to
RT	closely related bacteria.";
RL	Nature 423:81-86(2003).
DR	EMBL; AS017031; AAP26160.1; -.
DR	TIGR; B2289; -.
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPR002086; Aldehyde_dehydr.
DR	Pfam; PF00171; aldedh; 1.
KW	Complete proteome.
SQ	SEQUENCE 474 AA; 51998 MW; 70BC55347DC9D22D CRC64;
Query Match 30.8%; Score 736; DB 16; Length 474;	
Best Local Similarity 35.7%; Pred. No. 2.1e-37;	
Matches 168; Conservative 99; Mismatches 195; Indels 8; Gaps 5;	
QY	8 YNNGEWSNVNQIILLSPIDSSILGFPVPMTRREVDHAKGREALPAWALTVYRACY 67
DB	6 YINGQDKSVNTYKPLVAPSEETLAEIAQGTEDVEVAAKAKMKMTLSAYDRATI 65
QY	68 LKAAADIIRDEKEATVLAKEISKAYNASVTEVTRADLIRYAAEGRILSTADEGK 127
DB	66 LEVACKMDERREFAEIIAKEAKPIRAARGEVDRTVQYKFAAEAKRI---YGETLP 122
QY	128 MDASTG--HKLAVIRROPVIGVIAIAPYVNVLSGKIAPALIGGVNMFKPTQGSVS 185
DB	123 LQAAPGADRIAYTRKPIGVIGAITPENPLNVAKVGPALAGNTVVKPADQTPUS 182
QY	186 GLVAKAFAGLPAQVNTITGRGEIGDYIVHEEVNFINTGSTPVGQRIKLAGNR 245
DB	183 SYALIELEFEEAGLPKGMALIIISPGSTVGEALVNDVVASITFTGSPKVGIGAKAGLK 242
QY	246 PIMLELGGKDGAGIVLADADL--DNAAKQIVAGVDYSGQRTAKRVLVVEEVADEAEKI 304
DB	243 RVTELGSSNAAVIIDEVELTDELIERVKVGFVNGQVCIISVQRFVHEERDDFLSKL 302
QY	305 SENVAKLSVGP- PDNATVTPVIDNSADFTESLVDARQKAGEL-NEPKRDRGLLTGP 362
DB	303 RKAMETVWVGDFLEEDVSALISKVVERIDNNVWGEAKGANVLYCGNKRDERVFEPT 362
QY	363 LFDHVTLDKMLAWPEPGPIILPIIRVKDAEEAIAANKSDPGLQSSVFTDRFOKAFDIAN 422
DB	363 VLINVEHVSVQCOEVEGPLMTVNTFKFEDEAIEOVNNSRYGLQAGVFTNNLFKAMRAID 422
QY	423 KLEVGTVHNNKTRGPDNPFPLGLKSGAGVQGIKRYSTIEMTVKSVL 472
DB	423 ELEVGWGVINDIPTFRVDHMPYGVKSGTGREGIKYAEEMTKLVCI 472
RESULT 35	
Q9KC36	PRELIMINARY; PRT; 482 AA.
AC	Q9KC36;
DT	01-OCT-2000 (TREMELrel. 15, Created)
DT	01-OCT-2000 (TREMELrel. 15, Last sequence update)

GN	RA2289.
OS	Bacillus anthracis (strain Ames).
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=198094;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22608414; PubMed=12721629;
RA	Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA	Nelson K.E., Tettelin H., Pouts D.E., Eisen J.A., Gill S.R.,
RA	Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA	Kolman J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA	DeBoy R.T., Madupu R., Dougherty S.C., Durkin A.S., Haft D.H.,
RA	Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA	Berry J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA	Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA	Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.B.,
RA	Fraser C.M.;
RT	"The genome sequence of Bacillus anthracis Ames and comparison to
RT	closely related bacteria.";
RL	Nature 423:81-86(2003).
DR	EMBL; AS017031; AAP26160.1; -.
DR	TIGR; B2289; -.
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPR002086; Aldehyde_dehydr.
DR	Pfam; PF00171; aldedh; 1.
KW	Complete proteome.
SQ	SEQUENCE 474 AA; 51998 MW; 70BC55347DC9D22D CRC64;
Query Match 30.8%; Score 736.5; DB 2; Length 491;	
Best Local Similarity 35.8%; Pred. No. 2.1e-37;	
Matches 171; Conservative 84; Mismatches 197; Indels 25; Gaps 7;	
QY	5 YQY----YNGEHWK--SSVNVQIILLSPIDSSILGFPVPMTRREVDHAKGREALPAWAA 58
DB	8 YENMSLQIAGVRVAGKRDLDVDPFTQELKQIPLANREDIYAYSAQOAVAA 67
QY	59 LTVYRQYLKKAADIIRDEKEATVLAKEISKAYNASVTEVTRADLIRYAAEGRIL 118
DB	68 CGPSRAQVNLNVRIFDERDEIIDWIRESGS-----TRIKAQIENGRAALITQ 118
QY	119 ST----SADGKKDASTGHKLAVIRROPVIGVIAIAPYVNVLSGKIAPALIGGVN 174
DB	119 ESASLPSRVHGRILASDVPGKESRYREPLGVIGIISPNFPLHLTARSIALPALGNAC 178
QY	175 MKPTQGSVS-GLVAKAFAGLPAQVNTITGRGEIGDYIVHEEVNFINTGSTP 233
DB	179 VIKPASDTPVTGELLAHIFEAGLPKGLSVVVGSGSIGDAPVEHEVGFISTGSTQ 238
QY	234 VGORIGKLAG---MRPIMLELGGKDGAGIVLADADLDNAKQIVAGVDYSGQRTAKR 289
DB	239 VERNIGRTAGGEHLKHEVALELGNSPPVVLADADLDQAVNAAVVGFHLQGIQWALNR 298
QY	290 VLVVEEVADEAEKISENVAKLSVGPDPDNATV--PVDDNSADFTESLVDARQKAKE 348
DB	299 IIVDSVYDEFVNRKRYAERVKSLPYGDPSPKPTVGVGVINAKQLAGLQDKIATAKSEGARV 358
QY	349 LNEFKDGLLTGCLPDHVTLDKMLAWPEPGPIILPIIRVKDAEEAIAANKSDPGLQSS 408
DB	359 MVEGAQGNVLPVFPVADVTADMEIAREEIFGLVIGIQARDEAHALANSSEYGLSSA 418
QY	409 VFTDFQKAFDIANKLEVGTVHNNKTRGPDNPFPLGLKSGAGVQGIKRYSTIEMTV 465
DB	419 VFTSSLERGVKFAFGIBAGWTHINDIPVNDPNAPFGGKNSGLGRNGDWATEEFT 475
RESULT 34	
Q81QX6	PRELIMINARY; PRT; 474 AA.
AC	Q81QX6;
DT	01-JUN-2003 (TREMELrel. 24, Created)
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT	01-OCT-2003 (TREMELrel. 25, Last sequence update)
DE	Aldehyde dehydrogenase family protein.

01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
NADP-dependent glyceraldehyde-3-phosphate dehydrogenase.
GN BH1738.
OS Bacillus halodurans.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=86665;
EN [1]_TaxId=86665;
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
EA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT *Complete genome sequence of the alkaliphilic bacterium Bacillus
RI halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AF001513; BAB05457.1; -.
DS PIR: B83867; B83867.
DR HGSP: P51977; IBSX.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
KW Complete proteome.
SQ SEQUENCE 482 AA; 52424 MW; 29DFBCABEB1F8D4D CRC64;

Query Match 30.7%; Score 735; DB 16; Length 482;
Best Local Similarity 35.0%; Pred. No. 2.5e-37;
Matches 164; Conservative 90; Mismatches 209; Indels 6; Gaps 3

QY	9	VNGEKKSSVNCQIEILSPIDDSISLGFPVPAMTREVVDHMKAGREALPAWAALTVVERAQYL 68
DB	12	ISQWQHQQQSLEVIDEGTNHVIAITVPAATEADAKAATZAAKEGAEIARQMPTHERAAII 71
QY	69	HKAADIIFERDKEEIATVLAKETSKYNASVEVVRTLIRYAEAGSIEGLSTSADEGKM 128
DB	72	KRTAEIVTRDELFAQTIAKESKTIIEARSEVRRCITILTSGEAKRLHGHTIPFSOM 131
QY	129	DASTGH--KLAVIRQPVGVIATAPNYPNVNLGGSKIAPALIGGVNVWFPPPTQGSVSG 186
DB	132	---PGHRRRVGYFRFGPIIGIIVAITPFNDPLNVAHKIGPAIAGNSLIIKPSSTPLSA 188
QY	187	LVLAKAFENGAGVENITIGRSEIGDVIVHEHVFNFNTGSTPVQORIGKLAGMRP 246
DB	189	LRUVKALEAGLPKKIVQVITGHGVSIGPTLYKHDRVLLISFTGGYETGEKIARSAGVK 248
QY	247	IMLEGKGDAIGVLADLDNAAKQIVAGAYDYSGQRCTAIKRVLVVEEVADLAEKISE 306
DB	249	LAMELGSNSPIVLQDAELMEAVASCVSGCAHQNCICGVORIYYEQSVFNSTQSFVA 308
QY	307	NVAKLSVGDPFDNAT-VTPVDDNSADFIESLVDAKOKGAKEINFKRDQLLTTCGLEF 365
DB	309	QTKQLRLGDQSEETDIGMISEKEAKRIERWVEAKEEGARVLTGRGTGAYFEPTVLT 368
QY	366	HVTLDMLKAWEEPGLPIITRYKDAAEAVAANKSDFGLOSSVFFTRDPQKAFDIANKLE 425
DB	369	NVSPSRIAKEEAPAFIVTEGHSLTTEATARANDVDGFLQAGLFNNLTAFASAIEKLE 428
QY	426	VGVTHNNKTGRGDNFPFLGKSGAGVGOGIRYSIEAMTNVKSIVLDM 474
DB	429	VGGIWNVDSSVRIDAMPFGGIKKSGLGREGVRIALEEMTEQKVVAHL 477

RESULT 36
Q8TWC7
ID Q8TWC7 PRELIMINARY; PRT; 471 AA.
AC Q8TWC7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NAD-dependent aldehyde dehydrogenase.
GN MK108.
OS Methanopyrus kandleri.

GN GABD OR Z3959 OR ECS3522.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai D.J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Postbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 RA Khara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22 (2001).
 DR EMBL: AE005495; AAC57768.1; -;
 DR ENBL: AP025622; BAB36945.1; -;
 DR PIR: B91069; B91069.
 DR PIR: D85913; D85913.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 DR Complete proteome.
 KW SEQUENCE 482 AA; 51817 MW; 0CF3D4790A9EA015 CRC64;

Query Match 30.0%; Score 717; DB 16; Length 482;
 Best Local Similarity 33.9%; Pred. No. 3.3e-36;
 Matches 162; Conservative 92; Mismatches 208; Indels 16; Gaps 7;

QY 6 QNYVNGSWKSSVN--QIEILSPIDSSLGFPVPMTRFEEVDHAKGREALPAWALTVE 63
 DB 12 QALINGELDANNGEVIDVTPANGDKLSGVPKMGADETRAADANRALPWALTAK 71
 QY 64 RAQYLHKAADIETRKEELATVLAKELISKAYNASVTEVTRADLIRYAEIGIRLSTAD 123
 DB 72 RANLRWFWNLAMHQQDRLARLMTLEQKPLAEAKGEISYAAGFIEWFAEKGRIYDGI 131
 QY 124 EGGKMDASTGHKLAVIRROPVGLVLAIPYVPVNLSSKIAPALGNGVWVEFPPTQS 183
 DB 132 FGHQAD-----KRLIVKQIPGVTAITPNWFFAGMIRKAGALAGCTVWLPNSQTP 186
 QY 184 VSGLVLAFAEAGLPAGVFNITITGRSGEIGDYIVHEEVEVNFNFTGTPVGQRIKLAG 243
 DB 187 FSALALAEALIRAGIPAGVFNVTGSAGVGNELTSNPLRYKLSFTGSTEIGRLMEQCA 246
 QY 244 --MRPIMLEGGKDGAGIVADADLDAAKQIVAGAYDSGQRTAKRVLVVEEVADELA 301
 DB 247 KDIKVSELEGGNAPFIVFDADLDKAVEGALSKFRNAGTCVCANRLYYQDGVYDRFA 306
 QY 302 EKISENVAKLSVGDPFD-NATVTPVDDNSADFIESLVVDARQKAKEL---NEFKDGR 357
 DB 307 EKLQQAQVSKLHGGGLDGVITGLIDEKAVAKVEEHIADALEKARVVCCKAHERGN 366
 QY 358 LLTGLPDHVTLDMLKAWEEPFGPILPIIRVKDAEEVAIAANKSDFGLQSSVTRDFOKA 417
 DB 367 FFQPTILVDYANAKVSKKEETFGFLAPFRFKDADVIAQANDTEFLAANYFYARDLSRV 426

QY 418 FDIANKLEVGTVHINKTG-RGDNFPELGLKSGAGVGGVIRYSTEAMTVKSVILDM 474
 DB 427 FRVGEALYGVIGVGN--TGLISNEVAPPGIKASGLREGSGYGLIEDYLEIKYMCIGL 482
 RESULT 42
 ID Q9HK01 PRELIMINARY; PRT; 493 AA.
 AC Q9HK01;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Probable aldehyde dehydrogenase.
 GN TA0809.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatiales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 acidophilum";
 RL Nature 407:508-513 (2000).
 DR EMBL: AL445065; CAC11938.1; -;
 DR HSSP: P05091; ICW3.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 DR Complete proteome.
 KW SEQUENCE 493 AA; 54782 MW; 2F6A79DD96BE7FF5 CRC64;

Query Match 29.9%; Score 716.5; DB 17; Length 493;
 Best Local Similarity 34.8%; Pred. No. 3.6e-36;
 Matches 167; Conservative 97; Mismatches 191; Indels 27; Gaps 9;

QY 8 YNNGEW--KSNVQIEILSPIDSSLGFPVPMTRFEEVDHAKGREALPAWALTVE 65
 DB 6 YIDGQWNNSSGKTVKYSPTVGVIGRFEATRDVDRAIDAAEDAFWANDLGSVERS 65
 QY 66 QYLHKAADIETRKEELATVLAKELISKAYNASVTEVTRADLIRYAEIGIRLSTADEG 125
 DB 66 XIIRYAKELIEKRAELENIIMEENGKPVKEAKEYVDGVIDQIYYAEAWARKINGEVVEG 125
 QY 126 GKMDASTGHKLAVIRROPVGLVLAIPYVPVNLSSKIAPALGNGVWVEFP--PTGS 183
 DB 126 -----TSSHRKIYQYKVPVGVIVALTPNWFFAGMIRKAPALGTGNTVLPSSDTPGS 180
 QY 184 VSGLVLAFAEAGLPAGVFNITITGRSGEIGDYIVHEEVEVNFNFTGTPVGQRIK--L 241
 DB 181 AEWIV--RKFEVGEVKGVLNFTITGRSGEIGDYIVHEEVEVNFNFTGSTATQRIQKAS 238
 QY 242 AQMRPIMLEGGKDGAGIVADADLDAAKQIVAGAYDSGQRTAKRVLVVEEVADELA 301
 DB 239 ANMAKLELGGKAPPMWVKQADMDNALKTLWAKYWNAGQSCIAAERLYVHEDYDTFM 298
 QY 302 EKISENVAKLSVGDPFDNATVTPVDDNSADFIESLVVDARQKAK-----ELNEFK 353
 DB 299 SFVELSRKLAGDGP-KRADMGPLINKGALQATSEIVEAKESGAKILFGSQSLSGPY 357
 QY 354 RDGRLLTGLDHTVLDMLKAWEEPFGPILPIIRVKDAEEVAIAANKSDFGLQSSVTRD 413
 DB 358 RNYFFELPTIIGNAQKSKIQQEELIPAPVIGARKISSVEEYDNLANDSKYGLASVLTGD 417
 QY 414 FOKAFDIANKLEVGTVHINKTG-RGDNFPELGLKSGAGVGGVIRYSTEAMTVKSVILDM 471
 DB 418 PNIIFEASERIRFGLYVNPMP--GPEASQGYHTGFRMTQAGSGSKYGISLYLKNII 474

RESULT 47

Q81IM7 PRELIMINARY; PRT; 483 AA.
 AC 081IM7
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Succinate-semialdehyde dehydrogenase (EC 1.2.1.16).
 GN BC0357.
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=226900;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22608415; PubMed=1271630;
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
 RA Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
 RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
 RA Grechkin Y., Pusch G., Hasekorn R., Fomstein M., Ehrlich S.D.,
 RA Overbeek R., Kyrpides N.;
 RT "Genome sequence of Bacillus cereus and comparative analysis with
 RT Bacillus anthracis.";
 RL Nature 423:87-91(2003).
 DR EMBL; AE016999; AP07397.1; -.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0009013; F:succinate-semialdehyde dehydrogenase [NAD(P) . .]; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 483 AA; 52259 MW; 3815993C33C05893 CRC64;

Query Match 29.7%; Score 711.5; DB 16; Length 483;
 Best Local Similarity 35.3%; Pred. No. 7.2e-36;

Matches 169; Conservative 83; Mismatches 204; Indels 23; Gaps 8;

QY 8 YNGEWSVQIIEILSPIDSSGLFVPAMTREVDPHAMKAGREALPAWAAALTVYERAY 67
 DB 16 YINGEWITLQEIENWNPATKEIFATVPKGVTEAKQAVDAHAFAKTSKLTAADRAIK 75
 QY 68 LHKAADIERKEIAIVLAKELISKAYNASVTEVVTADLIRYAABEGIRLSTSADEGK 127
 DB 76 LKWFLLIDENKEIAIMTKQKGFPAEALGEVYNVANSFVWEAEGKRV-----YGE 130
 QY 128 MDASTGHKLAVIRQPVGIVLAIPYNPVNLGSKIAPALIGNVNKPPTQGSVSL 187
 DB 131 IPAAHKNRILVMKQPGVWAAITWNPFAAMTRKVAALAGCTAVVPASQPTLAL 190
 QY 188 VLAKAPAGAGLPAGVNTTGRSEIGDYIVVEHEVNFINTGSTPVQRIKLAG--MR 245
 DB 191 KLAEALAEADIPKGVINIITGSAKAIAADTWEDGRVRKVSFTGSTETGKELMASAAQTMK 250
 QY 246 FIMLELGGKAGIVLADADLNAKQIVAGAYDSQRCATKRVLVVEVVADELAELKIS 305
 DB 251 KVSLEIGHAPFIVMDADLNAKQIVAGAYDSQRCATKRVLVVEVVADELAELKIS 310
 QY 306 ENVAKLSVGDPF--DNATVTPVDDNSADFIESLVVDARQKAKELNEFEKRA---DGLLT 360
 DB 311 KAVGQLKVGDFGDTGTVGLDENAVSKVQEHIEDAIQKGTVLVYGQKVAELSGHTIQ 370
 QY 361 P---GLFDHVTLMKLAWEPPGPILPIIRVKDAEEAVATANKSDFGQSSVFTDFOKAF 417
 DB 371 PTVIGLANDTLMCN---BETFGVPVAKFKTVDEVERANNTPYGLAAVIFTKDISQA 427
 QY 418 FDIANKLEVGVTHNKTGRGP--NPFPLGLKSGAGVQIRYSIEAMTNKSVILDM 474
 DB 428 FQISEALEYGIIGNDGL---PSVAQAPFGFKESGIGREGHFGIEYLEIKYISLGL 483

RESULT 48

Q8ELI8 PRELIMINARY; PRT; 475 AA.
 AC 08ELI8
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Aldehyde dehydrogenase (Retinol).
 GN ALDY1 OR VNG2513G.

Q8ELI8;
 AC 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Aldehyde dehydrogenase.
 GN OB3239.
 OS Oceanobacillus iheyensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-HT831 / DSM 14371 / JCM 11309;
 RC MEDLINE=2220767; PubMed=12235376;
 RT Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments.";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 DR EMBL; AP004604; BAC15195.1; -.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 475 AA; 51562 MW; DD0C6A55752739EC CRC64;

Query Match 29.7%; Score 711; DB 16; Length 475;
 Best Local Similarity 33.1%; Pred. No. 7.5e-36;
 Matches 156; Conservative 105; Mismatches 207; Indels 4; Gaps 4;

QY 3 KEYQNVNGEWSVQIIEILSPIDSSGLFVPAMTREVDPHAMKAGREALPAWAAALTV 61
 DB 4 RKAQSFINGEWITLQEIENWNPATKEIFATVPKGVTEAKQAVDAHAFAKTSKLTAAD 63
 QY 62 YERAQVLHKAADIERDKEEIAIVLAKELISKAYNASVTEVVTADLIRYAABEGIRLST 121
 DB 64 IERSKILKKAAYLLENKEKFAKLSSELGKPLKDTLGEVDRSIEFLELSGEEAKRLHGE 123
 QY 122 ADEGRKMDASTGHKLAVIRQPVGIVLAIPYNPVNLGSKIAPALIGNVNKPPTQ 181
 DB 124 TIPGSSSERGL--NTIASIYRVPVGVAAITPFAALNVLCHVKVGFAGGNTVLLKPA 182
 QY 182 GVSGLVLAKAFAPAGVNTTGRSEIGDYIVVEHEVNFINTGSTPVQRIKGL 241
 DB 183 TLLIATAFQLLFAGPKNIAEHLIG--GVVGQOIVKDERINVISFTGGLVASRNI 241
 QY 242 AGMRPIMLELGGKAGIVLADADLNAKQIVAGAYDSQRCATKRVLVVEVVADELA 301
 DB 242 AGIKTFLFELGNAATVHQDADIETAAALCAKCTGYSNQSQCISQIRIYVQEEIKDN 301
 QY 302 EKISENVAKLSVGDPF--DNATVTPVDDNSADFIESLVVDARQKAKELNEFEKRLT 360
 DB 302 ELKQEVSKLVGDFLPHDTIGTIVDEKAAGRVIEINEAVESGAETIQQTKNGASVE 361
 QY 361 PGLFDHVTLMKLAWEPPGPILPIIRVKDAEEAVATANKSDFGQSSVFTDFOKAF 420
 DB 362 PTVLFNPPKSKVCEEVFGPIVSVITPKNIEEAIQESNDSFGLTGTGIFTNQIDLA 421
 QY 421 ANKLEVGVTHNKTGRGP--NPFPLGLKSGAGVQIRYSIEAMTNKSVIL 472
 DB 422 ADEMEVGGIVGTSNFDLPHLPYGVGVNKGIGREGPRVLAEBEMTESQWVL 473

RESULT 49

Q9HMJ6 PRELIMINARY; PRT; 489 AA.
 AC Q9HMJ6
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Aldehyde dehydrogenase (Retinol).
 GN ALDY1 OR VNG2513G.

OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 OC Archaea: Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahaitas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Laeky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR ENBL; AB005128; AAG20575.1; -;
 DR PIR; C84401; C84401.
 DR HGSP; P51977; 1BX5.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 DR Complete proteome.
 KW Complete proteome.
 SQ SEQUENCE 499 AA; 52022 MW; 142C3F4D8CFB126F CRC64;

Query Match 29.7%; Score 710.5; DB 17; Length 489;
 Best Local Similarity 35.9%; Pred. No. 8.5e-36;
 Matches 167; Conservative 84; Mismatches 199; Indels 15; Gaps 6;
 YQ 8 YVNGEW--KSVNQIEILSPIDSSLGFGPAMTREVVDHAKAGREALPAAALTVYERA 65
 DB 15 YIDGYPVPGDGTLEVDPPSTRELTFTEIPAGTDDVDAEAAQAQEWADTFPQERA 74
 YQ 66 QYLHKAADIITIERKEEATVLAKEISKAYNASVTEVTRTADLIRYAAEGIRLSTSADEG 125
 DB 75 RIVAAASEILASHRDEVDLLATESGTYTKGAEEFAESVGITDEAASEPTFMN-----G 129
 YQ 126 GMDASTGCHKLAVIRQPVGIVLAIPYVNLGSKIAIPALIGNVNMFKEPTQGSVS 185
 DB 130 DHKQSVVEGKENTVVRQPGVGVVGIISPNFPLNLSIRAVAPAAAGNSVVLKASNTPTVT 189
 YQ 186 -GLVLAKAPAEAGLPAGVNTITGSGEIGDYVEHEVNFNFTGTPVGORIGKLAGM 244
 DB 190 GGLLAKVFEAGLPAGVNVVTGSGEIGDRIAGHPHADVISFTGSLAGQVAATAGE 249
 YQ 245 RPIM--LELGKDGAGIVLADADLNAAKQIVAGYDYGQRCCTAIRVLVVEEVADEIAE 302
 DB 250 NLALPAMELGNNVHIVTEDADLDDAVDSGVFGSLHGQICISINRHVVHEDVYDEVE 309
 YQ 303 KISENVAKLSVGDPPDNATV--PVIDNSADFIESLVVDARQKAGKELNPKRGELLTP 361
 DB 310 RLTERAALPTGTAHDADTVGPIIDESQRDQILDIERTVDAGATLETGGDHGLVVEP 369
 YQ 362 GLFDHVTLDMLKAWBPFGPILPIIRVKQAEAAVIAKNSDFGLQSSVTRDFQKADIA 421
 DB 370 TVLSDATNDMAAACNEHGFVAPVIPSDDDEAVIANSTYGLSGSVHAGTGDGARDIA 429
 YQ 422 NKLEVGTVINNTKRGDNPFFFLGLKSGAGVQGIYRYSIEAMTN 466
 DB 430 DRIDTGMHINDQVNVQEPHLPFGYKNSGLG----RYNGEYVLN 470

RESULT 50

Q8YQP9 PRELIMINARY; PRT; 498 AA.
 AC Q8YQP9;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Aldehyde dehydrogenase.

GN ALR3771.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120";
 RL DNA Res. 8:205-213(2001).
 DR ENBL; AF003594; BAB75470.1; -;
 DR PIR; AD2277; AD2277.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 498 AA; 54302 MW; 8FE079D0941B9837 CRC64;

Query Match 29.7%; Score 710.5; DB 16; Length 498;
 Best Local Similarity 33.0%; Pred. No. 8.7e-36;
 Matches 159; Conservative 105; Mismatches 199; Indels 19; Gaps 8;
 YQ 6 QYVNGEWKSVNQIEILS--PIDSSLGFGPAMTREVVDHAKAGREALPAAALTVY 62
 DB 8 QYVINGQWLANAATETTLNHNPAKSEIVATFPRSQADVDVRAAARQAQYSGWRKVPAP 67
 YQ 63 ERAQYLHKAADIITIERKEEATVLAKEISKAYNASVTEVTRTADLIRYAAEGIRLSTSA 122
 DB 68 ARAEYIFRVGSELLQHKELAQILSRMGKPLTEARGVQGVQVCAYFSAEGRLLF--- 124
 YQ 123 DEGGKMDASTGCHKLAVIRQPVGIVLAIPYVNLGSKIAIPALIGNVNMFKEPTQG 182
 DB 125 --GQYTPSEMKNFAMTMRPIGVCALITPNFPVAIPCWKAMPALVCGNTVLKPAEDT 182
 YQ 183 SVSGLVLAKAPAEAGLPAGVNTITGSGEIGDYVEHEVNFNFTGTPVGORIGKLA 242
 DB 183 PACATKLTETIPAAAGLPAGVNLVHGVEGVKALVEHPENIDLVSFTGSSATGAYVETC 242
 YQ 243 GM--RPIMLELGKDGAGIVLADADLNAAKQIVAGYDYGQRCCTAIRVLVVEEVADEL 300
 DB 243 GRTHKRVCCLEMGKNAQVVDADLDELALDGLWAGAGTGTGORTATSLILLHROIKEKF 302
 YQ 301 ABKISENVAKLSVGDPPDNAT--VTFVIDNSADFIESLVVDARQKAGKEL--NEFKRGR 357
 DB 303 TTMLEKRTSQLRGAGCTETPDIGIINNQLQORVHEYNIAEEGAKILIGGEIATEGQ 362
 YQ 358 L-----LTFGLFDRHVTLDMLKAWBPFGPILPIIRVKQAEAAVIAKNSDFGLQSSVTR 412
 DB 363 LQOGYFFPTILDVNPQVRVAREEIFGPVVALIEVSTFEAAIALNDTKYGLSSSVTR 422
 YQ 413 DFQKAFDIANKLEVGTVINNTKRGDNPFFFLGLKSGAG--VOGIRYSIEAMTNVKSIV 471
 DB 423 DINRAVFAMRDIEAGITYINGFTTGAEVHLPFGGVKQVQNGHREAGTALDVTFTKSVY 482
 YQ 472 LD 473
 DB 483 VD 484

Search completed: April 13, 2004, 14:22:27
 Job time : 51 secs